



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

Μελέτη της Κοινωνικής Ηλικίας της Πληροφορίας: Μια Πραγματική Περίπτωση Χρήσης στο Facebook

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΔΗΜΗΤΡΙΟΣ Α. ΜΠΑΣΙΝΑΣ

Επιβλέπων: Συμεών Παπαβασιλείου
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2022



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

Μελέτη της Κοινωνικής Ηλικίας της Πληροφορίας: Μια Πραγματική Περίπτωση Χρήσης στο Facebook

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΔΗΜΗΤΡΙΟΣ Α. ΜΠΑΣΙΝΑΣ

Επιβλέπων: Συμεών Παπαβασιλείου
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 16^η Μαρτίου 2022.

.....

Συμεών Παπαβασιλείου

Καθηγητής Ε.Μ.Π.

.....

Ιωάννα Ρουσσάκη

Επίκουρος Καθηγήτρια Ε.Μ.Π.

.....

Βασίλειος Καρυώτης

Αναπληρωτής Καθηγητής
Τμήμα Πληροφορικής Ιόνιο
Πανεπιστήμιο

Αθήνα, Μάρτιος 2022

.....
ΔΗΜΗΤΡΙΟΣ Α. ΜΠΑΣΙΝΑΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Δημήτριος Μπασινάς, 2022

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η αυξημένη μεταφορά πληροφορίας, μέσω των κοινωνικών - και όχι μόνο - δικτύων, καθώς και η χρησιμότητα της για τον προορισμό, έχει δημιουργήσει την ανάγκη για τον προσδιορισμό του χρόνου από τη δημιουργία της με σκοπό την απόκτηση της πιο πρόσφατης. Στην παρούσα διπλωματική εργασία ερευνάται η ηλικία της πληροφορίας (Age of Information, γνωστή ως AoI) στο πλαίσιο των κοινωνικών δικτύων, μέσω των σχολίων μίας δημοσίευσης στο Facebook. Αναπτύσσεται και υλοποιείται η ιδέα της χρήσης της κοινωνικής AoI, ως το χρόνο από το τελευταίο σχόλιο, για την εύρεση της αξίας της ανανέωσης της σελίδας μιας συγκεκριμένης δημοσίευσης. Αφού προβάλλεται ο τρόπος άντλησης τέτοιων πληροφοριών από κοινωνικά δίκτυα μέσω των διεπαφών προγραμματισμού εφαρμογών (APIs), γίνεται μελέτη της κατανομής που ταιριάζει στις μέγιστες τιμές της ηλικίας, βάσει της οποίας εισάγεται ένα μοντέλο που ποσοτικοποιεί την προαναφερθείσα αξία. Παρουσιάζεται η πορεία της τιμής της αξίας με τη μεταβολή ορισμένων παραμέτρων και εξετάζεται η ποσοτική και ποιοτική συσχέτιση της με την κοινωνική AoI. Γίνεται σύγκριση μεταξύ των αποτελεσμάτων του μοντέλου και μιας προσομοίωσης πάνω στα ακριβή δεδομένα, η οποία δείχνει ικανοποιητική προσέγγιση της αξίας. Τέλος, προτείνεται η επέκταση της μελέτης για την εύρεση βέλτιστης συχνότητας πρόσβασης των χρηστών στα κοινωνικά δίκτυα με σκοπό την αντιμετώπιση της συμφόρησης στο δίκτυο, που αν επιτευχθεί, θα οδηγήσει σε μείωση της ηλικίας της πληροφορίας.

Λέξεις – κλειδιά: ηλικία της πληροφορίας, κοινωνικά δίκτυα, αξία ανανέωσης, Facebook, κατανομή δεδομένων

Abstract

The increased transfer of information, through social - and not only - networks, as well as its usefulness for the destination, has created the need to determine the time since its creation in order to obtain the most recent. The current diplomatic work explores the age of information (Age of Information, known as AoI) in the context of social networks, through the comments of a Facebook post. The idea of using social AoI, as the time since the last comment, is developed and implemented in order to find the value of refresh of the page of a particular post. Once the way to derive such information from social networks through application programming interfaces (APIs) is viewed, the distribution fitted by peak age values is being studied, on the basis of which a model is introduced to quantify the above value. The evolution of the value price with the change of certain parameters is presented and its quantitative and qualitative correlation with social AoI is examined. A comparison is made between the results of the model and a simulation on the exact data showing a satisfactory approximation of value. Finally, it is proposed to extend the study to find an optimal frequency of user access to social networks so as to address network congestion, which, if achieved, would lead to a reduction in the age of information.

Keywords: age of information, social networks, value of refresh, Facebook, data distribution

Ευχαριστίες

Με την παρούσα διπλωματική εργασία ολοκληρώνονται οι σπουδές μου στην Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Αρχικά, θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Συμεών Παπαβασιλείου, για την εμπιστοσύνη που μου έδειξε με την ανάθεση της διπλωματικής εργασίας και την άριστη συνεργασία κατά τη διάρκεια εκπόνησης της. Εν συνεχεία, θα ήθελα να εκφράσω την ιδιαίτερη ευγνωμοσύνη μου στον Αναπληρωτή Καθηγητή του Τμήματος Πληροφορικής του Ιόνιου Παν/μίου κ. Βασίλειο Καρυώτη για τη συνεχή καθοδήγηση και βοήθεια που μου παρείχε ώστε να υλοποιηθεί η παρούσα εργασία.

Τέλος, ευχαριστώ θερμά την οικογένεια μου για τα εφόδια που μου εξασφάλισε και τη στήριξη που μου προσέφερε καθ' όλη τη διάρκεια των σπουδών μου, αλλά και τους φίλους μου οι οποίοι στάθηκαν δίπλα μου και έκαναν τη φοιτητική μου πορεία ομορφότερη.

Μπασινάς Δημήτριος

Αθήνα, Μάρτιος 2022

Περιεχόμενα

Περίληψη	1
Abstract	2
Ευχαριστίες	3
Περιεχόμενα	4
Ευρετήριο Σχημάτων	7
Ευρετήριο Πινάκων	10
Κεφάλαιο 1: Εισαγωγή	11
1.1 Εισαγωγή στο Αντικείμενο της Διπλωματικής Εργασίας	11
1.2 Αντικείμενο της Διπλωματικής Εργασίας	11
1.3 Οργάνωση της Διπλωματικής Εργασίας	12
Κεφάλαιο 2: Βασικά Στοιχεία AoI	13
2.1 Ορισμός AoI	13
2.2 Μετρικές της AoI	15
2.2.1 Average Age	15
2.2.2 Peak Age και Average Peak Age	16
2.3 Η AoI σε Σχέση με Άλλες Παραμέτρους	17
Κεφάλαιο 3: Μοντέλα AoI Συνεχούς Χρόνου	20
3.1 Μοντέλα με 1 Εξυπηρετητή	20
3.1.1 M/M/1 - FIRST COME FIRST SERVED	20
3.1.2 M/D/1 - FIRST COME FIRST SERVED	22
3.1.3 D/M/1 - FIRST COME FIRST SERVED	23
3.1.4 Σύγκριση M/M/1, M/D/1, D/M/1	23
3.2 Μοντέλα με Περισσότερους Εξυπηρετητές	24
3.2.1 Informative – Obsolete Πακέτα	25
3.2.2 M/M/∞	25
3.2.3 M/M/2	27
Κεφάλαιο 4: Πολιτικές και Τεχνικές Μείωσης AoI	29
4.1 LCFS Πολιτική	29
4.1.1 LCFS Without Preemption`	29
4.1.2 LCFS With Preemption	31

4.1.3 Σύγκριση FCFS, LCFS	32
4.2 Διαχείριση Πακέτων	34
4.2.1 Ανάλυση Διαχείρισης Πακέτων	34
4.2.2 Σύγκριση M/M/1/2* και M/M/1	37
4.2.3 Σύγκριση M/M/1/2* και M/M/∞	38
4.3 Σύγκριση Διαχείρισης Πακέτων και LCFS	39
4.4. Just-In-Time Πολιτική	39
Κεφάλαιο 5: Μοντέλα AoI Διακριτού Χρόνου	43
5.1 Εισαγωγή στα Διακριτά Μοντέλα	43
5.2 First Come First Served Ουρές	44
5.2.1 G/M/1 – First Come First Served	45
5.2.2 M/G/1 – First Come First Served	45
5.2.3 BER/G/1 – First Come First Served	46
5.2.4 BER/G/1 with vacations – First Come First Served	46
5.2.5 G/BER/1 – First Come First Served	47
5.3 Συσχέτιση Διακριτής με Συνεχή AoI	48
5.3.1 Σύγκριση Συνεχούς M/M/1 με Διακριτή BER/BER/1	49
Κεφάλαιο 6: Age of Information στα Κοινωνικά Δίκτυα και APIs	51
6.1 Κοινωνικά Δίκτυα	51
6.2 Κοινωνική Ηλικία της Πληροφορίας (Social Age of Information)	51
6.3 APIs	51
6.3.1 Εισαγωγή στα APIs	52
6.3.2 Social Networks APIs	53
6.3.3 Facebook API	53
Κεφάλαιο 7: Παρουσίαση Dataset και Data Analysis	56
7.1 Άντληση και Επεξεργασία Δεδομένων από το Dataset	56
7.2 Εύρεση Social Age of Information	58
7.3 Fit της Peak Social Age of Information	59
7.3.1 Fitter	60
7.3.2 Distfit	62
Κεφάλαιο 8: Value of Refresh	66
8.1 Μοντέλο για τη Value of Refresh	66
8.2 Υπολογισμός VoR	68
8.3 Η Χρησιμότητα Περί Χρόνου	71
8.4 Σχέση Peak Social Age of Information με Value of Refresh	74
8.5 Προσομοίωση και Σύγκριση με το Μοντέλο	78

Κεφάλαιο 9: Σύνοψη	82
9.1 Συμπεράσματα	82
9.2 Μελλοντικές Επεκτάσεις	82
Βιβλιογραφία	84

Ευρετήριο Σχημάτων

Σχήμα 2.1 – Βασικό μοντέλο συστήματος.....	13
Σχήμα 2.2 – Παράδειγμα ενός sample path για ηλικία $\Delta(t)$	14
Σχήμα 2.3 – Χρονική απεικόνιση Age Of Information, Delay, Interdelivery Time.....	18
Σχήμα 2.4 – Η ηλικία της πληροφορίας και η καθυστέρηση σε συνάρτηση με το ρυθμό δειγματοληψίας για M/M/1 FIFO ουρά.....	19
Σχήμα 3.1 – Η μέση ηλικία της πληροφορίας για τα M/M/1, M/D/1, D/M/1 ως προς τη χρησιμοποίηση για $\mu=1$	24
Σχήμα 3.2 – Η ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το μοντέλο M/M/ ∞	25
Σχήμα 3.3 – Η ηλικία της πληροφορίας σε συνάρτηση με τη χρησιμοποίηση για διάφορες τιμές του μ για το μοντέλο M/M/ ∞	26
Σχήμα 3.4 – Η ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το μοντέλο M/M/2.....	27
Σχήμα 3.5 – Η ηλικία της πληροφορίας σε συνάρτηση με τη χρησιμοποίηση για τα μοντέλα M/M/2, M/M/1 για $\mu=1$	28
Σχήμα 4.1 – Παράδειγμα πορείας average AoI σε monitor με LCFS without preemption.....	30
Σχήμα 4.2 – Παράδειγμα πορείας average AoI με LCFS with preemption.....	31
Σχήμα 4.3 – Η ηλικία της πληροφορίας (σε έναν monitor) σε σχέση με τη χρησιμοποίηση του εξυπηρετητή για ένα μοντέλο M/M/1 FCFS,LCFS με και χωρίς δικαίωμα προτίμησης για $\mu=1$	33
Σχήμα 4.4 – Η μέση ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το μοντέλο M/M/1/1.....	34
Σχήμα 4.5 – Η μέση ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το μοντέλο M/M/1/2.....	35
Σχήμα 4.6 – Η μέση ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το μοντέλο M/M/1/2*.....	35
Σχήμα 4.7 – Η μέση ηλικία της πληροφορίας σε συνάρτηση με το ρυθμό αφίξεων λ για τα μοντέλα M/M/1/1, M/M/1/2, M/M/1/2* για σταθερό $\mu=1$	36

Σχήμα 4.8 – Η μέση ηλικία της πληροφορίας σε συνάρτηση με το ρυθμό αφίξεων λ για τα μοντέλα $M/M/1$, $M/M/1/2^*$ για σταθερό $\mu=1$	37
Σχήμα 4.9 – Η μέση ηλικία της πληροφορίας σε συνάρτηση με το ρυθμό αφίξεων λ για τα μοντέλα $M/M/\infty$, $M/M/1/2^*$ για σταθερό $\mu=1$	38
Σχήμα 4.10 – Μοντέλο συστήματος με ανάδραση.....	39
Σχήμα 4.11 – Το σύστημα με ACK για ανάδραση για zero-wait πολιτική.....	40
Σχήμα 4.12 – Η μέση ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το σύστημα του σχήματος 4.10 με πολιτική Just-In-Time.....	40
Σχήμα 4.13 – Διάγραμμα ροής για τη Zero-Wait πολιτική.....	41
Σχήμα 5.1 – Η ηλικία της πληροφορίας της διακριτής $A(t)$	43
Σχήμα 5.2 – Η μέση ηλικία της πληροφορίας σε συνάρτηση με το μέσο χρόνο διακοπών για το μοντέλο $BER/G/1$ με διακοπές για σταθερό $\mu=0.75$ και εξυπηρετήσεις γεωμετρικά κατανομημένες ($BER/Geo/1$). Η κατανομή του V μπορεί να είναι Γεωμετρική, Ομοιόμορφη και Ντετερμινιστική. Οι συνεχείς γραμμές αναπαριστούν πιθανότητα αφίξεως $\lambda=0.3$ ενώ οι διακεκομμένες $\lambda=0.6$	47
Σχήμα 5.3 – Η ηλικία της πληροφορίας της διακριτής $A(t)$ σε συνάρτηση με το χρόνο μαζί με την αντίστοιχη συνεχή διαδικασία $A_{cont}(t)$	49
Σχήμα 5.4 – Η μέγιστη και η μέση ηλικία της πληροφορίας σε σχέση με τη χρησιμοποίηση ρ για τη συνεχή $M/M/1$ και τη διακριτή $BER/BER/1$ για $\mu=0.8$...	50
Σχήμα 7.1 – Μέρος του dataset που επεξεργάστηκε και χρησιμοποιήθηκε.....	56
Σχήμα 7.2 – Η πορεία του αριθμού των comments στο συνολικό post σε σχέση με τις ημερομηνίες.....	57
Σχήμα 7.3 – Πλήθος comments στο post ανά ημέρα.....	57
Σχήμα 7.4 – Η πορεία του αριθμού των comments για τις 4 πρώτες ημέρες (α) ως προς τις ημερομηνίες και (β) ως προς το χρόνο.....	58
Σχήμα 7.5 – Η εξέλιξη της Social Age of Information σε σχέση με το χρόνο.....	58
Σχήμα 7.6 – Η εξέλιξη της Peak Social Age of Information σε σχέση με το χρόνο.	59
Σχήμα 7.7 – Το ιστόγραμμα των τιμών της Peak Social Age of Information.....	60
Σχήμα 7.8 – Fitting των κατανομών στο ιστόγραμμα των τιμών της Peak Social Age of Information.....	61
Σχήμα 7.9 – Αποτελέσματα του distfit με: (α) Απεικόνιση καλύτερων κατανομών με βάση το RSS και (β) Γραφικό fit της εκθετικής κατανομής στα δεδομένα.....	64
Σχήμα 8.1 – Η αναμενόμενη τιμή του $VoR (\bar{V})$ ως προς το μέσο ρυθμό μ των refreshes για $M = 1, 5, 10, 25, 50$	70

Σχήμα 8.2 – (α) Η αναμενόμενη τιμή του VoR (\bar{V}) ως προς το μέσο χρονικό διάστημα $1/\mu$ μεταξύ των refreshes για $M = 1, 5, 10, 25, 50$ και (β) Η αναμενόμενη τιμή του VoR (\bar{V}) ως προς τη μέση ποσότητα ρ νέων comments ανά refresh για $M = 1, 5, 10, 25, 50$ 71

Σχήμα 8.3 – (α) Η χρησιμότητα ανά δευτερόλεπτο ενός χρήστη ως προς το μέσο ρυθμό μ των refreshes για $M = 1, 5, 10, 25, 50$ και (β) Η χρησιμότητα ανά ώρα ενός χρήστη ως προς το μέσο ρυθμό μ των refreshes για $M = 1, 5, 10, 25, 50$ 72

Σχήμα 8.4 – Ο βέλτιστος μέσος ρυθμός των refreshes μ^* για τις διάφορες τιμές του μέγιστου αριθμού προβολής comments ανά refresh ($1 \leq M \leq 50$) 74

Σχήμα 8.5 – Η αναμενόμενη τιμή του VoR (\bar{V}) ως προς το μέσο ρυθμό λ δημοσίευσης comments για $\mu=0.0001$ και $M = 1, 5, 10, 25, 50$75

Σχήμα 8.6 – Η αναμενόμενη τιμή του VoR (\bar{V}) ως προς τη μέση peak Social Age of Information για $\mu=0.0001$ και $M = 1, 5, 10, 25, 50$ 76

Σχήμα 8.7 – (α) Η αναμενόμενη τιμή του VoR ως προς το μέσο ρυθμό μ των refreshes για $0.0001 \leq \mu \leq 0.001$ και $M = 25$, (β) Το σφάλμα της αναμενόμενης τιμής του VoR ως προς το μέσο ρυθμό μ των refreshes για $0.0001 \leq \mu \leq 0.001$ και $M = 25$, (γ) Η αναμενόμενη τιμή του VoR ως προς το μέσο ρυθμό μ των refreshes για $0.0001 \leq \mu \leq 0.001$ και $M = 50$, (δ) Το σφάλμα της αναμενόμενης τιμής του VoR ως προς το μέσο ρυθμό μ των refreshes για $0.0001 \leq \mu \leq 0.001$ και $M = 50$ 79

Ευρετήριο Πινάκων

Πίνακας 2.1 – Μέση AoI για διάφορες τιμές των delay, interdelivery time σε ουρά M/M/1 με $\mu=1$ και μεταβλητό λ	18
Πίνακας 7.1 – Οι κατανομές με τις αντίστοιχες τιμές των SSE.....	62
Πίνακας 7.2 – Οι κατανομές, οι RSS και οι παράμετροι τους.....	64
Πίνακας 8.1 – Οι μεταβλητές και η περιγραφή τους	68
Πίνακας 8.2 – Οι αναμενόμενες τιμές του VoR για διάφορες τιμές Average Peak Social Age of Information για $M=25$	77
Πίνακας 8.3 – Το μέσο Error για διάφορες τιμές του M και $0.0001 \leq \mu \leq 0.001$	80

Κεφάλαιο 1: Εισαγωγή

1.1 Πρόλογος

Με την ανάπτυξη της τεχνολογίας και ως εκ τούτου υπολογιστικών συστημάτων που επιτυγχάνουν υψηλές αποδόσεις, οι πληροφορίες που μεταφέρονται μεταξύ συσκευών μέσω των εφαρμογών είναι συνεχείς. Ειδικότερα, είτε η πληροφορία αλλάζει γρήγορα (π.χ. η τοποθεσία ενός αυτοκινήτου) είτε αλλάζει αργά (π.χ. η θερμοκρασία), η ανάγκη να είναι έγκυρη παραμένει αμετάβλητη. Δηλαδή, η πληροφορία έχει μεγαλύτερη αξία όσο πιο “φρέσκια” είναι. Επειδή οι παραδοσιακοί τρόποι μέτρησης της απόδοσης, όπως η καθυστέρηση, δεν μπορούν να χαρακτηρίσουν πλήρως το πόσο πρόσφατη είναι μια πληροφορία, παρουσιάστηκε η ανάγκη για τη δημιουργία ενός *metric* που θα μας το δείχνει με ακρίβεια και αποτελεσματικότητα. Έτσι, το 2011 προτάθηκε η έννοια της ηλικίας της πληροφορίας, Age of Information (AoI), από τους Sanjit Kaul, Roy Yates και Marco Gruteser [1] για να ποσοτικοποιηθεί η “φρεσκάδα” των γνώσεων που έχουμε για την κατάσταση ενός συστήματος. Με αυτόν τον τρόπο, παρουσιάστηκε και η ευκαιρία βελτίωσης της απόδοσης σε ένα -πραγματικού χρόνου- σύστημα. Για παράδειγμα, η γνώση της πραγματικής τοποθεσίας και ταχύτητας ενός οχήματος στην αυτόνομη οδήγηση ή της πραγματικής κατάστασης ενός καναλιού σε ένα σύστημα ασύρματης επικοινωνίας, είναι καίριας σημασίας και σκοπός είναι η βελτιστοποίηση της. Συγκεκριμένα, η AoI μετράει το χρόνο που έχει περάσει από την παραγωγή του τελευταίου πακέτου το οποίο έχει παραδοθεί με επιτυχία στον προορισμό. Γίνεται, λοιπόν, αντιληπτό πως μελετάται η ηλικία της πληροφορίας από την πλευρά του προορισμού (*destination-centric*), ο οποίος επιζητά να είναι όσο πιο επίκαιρη γίνεται αφού ένα ξεπερασμένο μήνυμα μπορεί να χάσει η αξία του.

1.2 Αντικείμενο της Διπλωματικής Εργασίας

Το γεγονός ότι μια πληροφορία είναι βασικό να φτάνει στον προορισμό έγκαιρα, ισχύει σε κάθε σύστημα. Στην παρούσα διπλωματική θα εξετάσουμε την ηλικία της πληροφορίας που μεταφέρεται μέσω των κοινωνικών δικτύων από τους χρήστες.

Προσαρμόζοντας την ηλικία της πληροφορίας - όπως παρουσιάζεται στο θεωρητικό υπόβαθρο παρακάτω - στο πλαίσιο των κοινωνικών δικτύων, μελετάται το πόσο πρόσφατη είναι η πληροφορία σε μία δημοσίευση στο Facebook με βάση τα σχόλια που γίνονται από διάφορους χρήστες. Βασιζόμενοι στο πώς κατανέμεται η κοινωνική ΑοΙ στη δημοσίευση αυτή, εισάγεται ένα μοντέλο που ποσοτικοποιεί την αξία της ανανέωσης του post από το χρήστη, για απόκτηση νέων πληροφοριών. Δεδομένου του ότι υπάρχουν πολλοί παράγοντες που επηρεάζουν την τιμή της αξίας αυτής ερευνάται η επίδραση του καθενός δίνοντας έμφαση στη συσχέτιση με την μέγιστη κοινωνική ηλικία της πληροφορίας. Παράλληλα, ελέγχεται η αντιστοίχιση των αποτελεσμάτων του μοντέλου με τις ακριβείς μετρήσεις ώστε να κριθεί η εφαρμογή του στα δεδομένα.

1.3 Διάρθρωση της Διπλωματικής Εργασίας

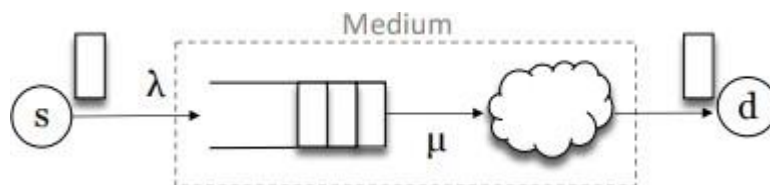
Στο πλαίσιο της διπλωματικής εργασίας, αρχικά, αναπτύσσεται το θεωρητικό υπόβαθρο (Κεφάλαια 2-5). Στο Κεφάλαιο 2 παρουσιάζονται κάποια δομικά στοιχεία της Age of Information με έμφαση σε ορισμένες μετρικές της. Στο Κεφάλαιο 3 εξετάζονται βασικά συνεχή μοντέλα τα οποία και συγκρίνονται μεταξύ τους ως προς τις προαναφερθείσες μετρικές. Στο Κεφάλαιο 4 περιγράφονται διάφοροι τρόποι και τεχνικές μείωσης της ΑοΙ με παράλληλη αποτίμηση της αποτελεσματικότητάς τους. Στο Κεφάλαιο 5 παρατίθενται βασικά μοντέλα διακριτού χρόνου και οι – ως προς την ΑοΙ - μετρικές τους συσχετίζονται με εκείνες των μοντέλων συνεχούς χρόνου. Στο Κεφάλαιο 6 εισάγεται η έννοια της κοινωνικής ηλικίας της πληροφορίας και επεξηγούνται τα API's προβάλλοντας έναν τρόπο εξαγωγής δεδομένων από το Facebook. Στο Κεφάλαιο 7 προβάλλεται το dataset που μελετήθηκε και γίνεται η επεξεργασία και η ανάλυση του. Στο Κεφάλαιο 8 παρουσιάζεται το μοντέλο που χρησιμοποιήθηκε και δίνονται τα αποτελέσματα της προσομοίωσης στο dataset σε σχέση με την κοινωνική Age of Information. Στο Κεφάλαιο 9 συνοψίζονται τα τελικά συμπεράσματα.

Κεφάλαιο 2: Βασικά Στοιχεία AoI

Στο Κεφάλαιο αυτό ορίζεται πλήρως η ηλικία της πληροφορίας ενώ παρατίθενται σημαντικές μετρικές της, όπως η μέση και η μέγιστη AoI. Παράλληλα, εξετάζεται η σχέση της με την καθυστέρηση και το χρόνο μεταξύ διαδοχικών παραλαβών των πακέτων.

2.1 Ορισμός AoI

Θεωρούμε ένα σύστημα με 2 κόμβους που είναι η πηγή και ο προορισμός. Μέσω μιας τυχαίας διαδικασίας $X(t)$ παράγονται δείγματα, τα οποία παρατηρούνται από την πηγή σε τυχαίες στιγμές και μεταφέρονται ως πακέτα στον προορισμό. Η μεταφορά αυτή γίνεται για να ανανεωθεί η τιμή της διαδικασίας στο κόμβο του προορισμού, γι' αυτό και τα μηνύματα αυτά λέγονται και ενημερώσεις κατάστασης (status updates). Κάθε status update i περιέχει πληροφορίες σχετικά με τις μεταβλητές που μας ενδιαφέρουν $X[t_i]$ καθώς και μια χρονοσφραγίδα t_i ώστε να αναγνωρίζεται η στιγμή που δημιουργήθηκε. Η ενημέρωση του προορισμού γίνεται ακριβώς με την άφιξη του πακέτου.



Σχήμα 2.1 – Βασικό μοντέλο συστήματος [9]

Θεωρώντας t_i , t'_i τις στιγμές που τα status updates πακέτα i δημιουργούνται στην πηγή και λαμβάνονται από τον προορισμό αντίστοιχα, μια τυχαία στιγμή t , ο δείκτης του πιο πρόσφατου status update θα είναι:

$$N(t) = \max \{i | t'_i \leq t\} \quad (2.1)$$

ενώ η χρονοσφραγίδα του:

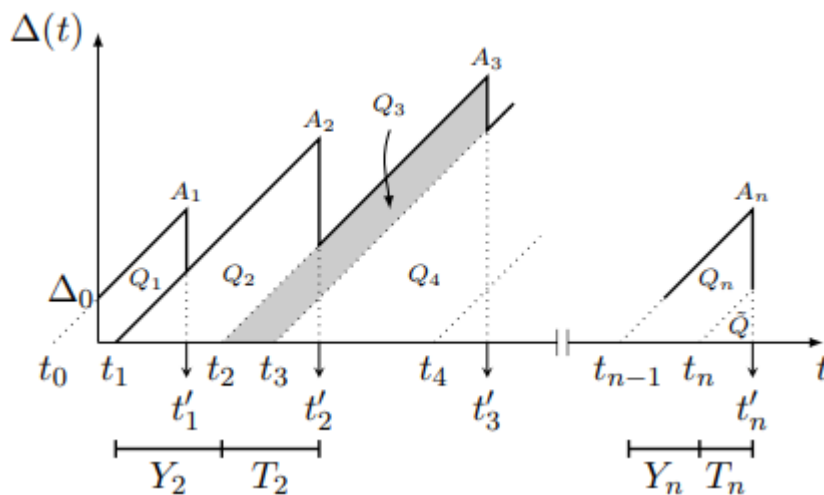
$$u(t) = t_{N(t)} \quad (2.2)$$

Έτσι, η ηλικία της πληροφορίας (Age of Information) ή status update age ή αλλιώς

απλά age ορίζεται ως η τυχαία διαδικασία:

$$\Delta(t) = t - u(t) \quad (2.3)$$

Στο παρακάτω διάγραμμα μπορούμε να δούμε ένα τυχαίο παράδειγμα για το πως εξελίσσεται η ΑοΙ σε συνάρτηση με το χρόνο.



Σχήμα 2.2 – Παράδειγμα ενός sample path για ηλικία $\Delta(t)$ [9]

Υποθέτουμε πως τη στιγμή $t=0$, ξεκινώντας την παρατήρηση του συστήματος, η ουρά είναι άδεια ενώ η ΑοΙ στον προορισμό είναι $\Delta(0) = \Delta_0$.

Η τυχαία μεταβλητή Y_i περιγράφει το χρόνο που μεσολαβεί μεταξύ δύο διαδοχικών αφίξεων πακέτων (interarrival time), δηλαδή των πακέτων $i-1, i$ και ορίζεται ως:

$$Y_i = t_i - t_{i-1} \quad (2.4)$$

ενώ η μεταβλητή T_i περιγράφει το χρόνο στο σύστημα για το πακέτο i (system time) και είναι:

$$T_i = t'_i - t_i \quad (2.5)$$

Βλέπουμε πως η ΑοΙ στον προορισμό, στα διαστήματα μεταξύ t'_{i-1} και t'_i , αυξάνεται γραμμικά με το χρόνο, αφού δεν υπάρχουν ενημερώσεις (updates) στον προορισμό,

λόγω απουσίας αφίξεων πακέτων σε αυτόν. Κατά τη λήψη ενός status update, βλέπουμε άμεση μείωση της AoI που στη συνέχεια θα αυξηθεί εκ νέου γραμμικά έως την άφιξη του επόμενου πακέτου στον προορισμό.

2.2 Μετρικές της AoI

Η ηλικία της πληροφορίας μπορεί να χαρακτηριστεί από δύο μετρικές. Αυτές είναι η μέση ηλικία (average age) και η μέγιστη ηλικία (peak age) [2], [3]. Οι μετρικές αυτές, είναι δυνατόν να χρησιμοποιηθούν αντί για την AoI, έχοντας η κάθε μια ξεχωριστά διαφορετικές χρήσεις και πλεονεκτήματα.

2.2.1 Average Age

Υποθέτουμε εργοδική τη διαδικασία $\Delta(t)$ ώστε να θεωρήσουμε τη μέση ηλικία ως προς το χρόνο (time average age). Παίρνουμε ως διάστημα παρατήρησης το $(0, \tau)$ και η time average age είναι:

$$\Delta\tau = \frac{1}{\tau} \int_0^{\tau} \Delta(t) dt \quad (2.6)$$

Το παραπάνω ολοκλήρωμα μπορεί να ερμηνεύσει και την περιοχή κάτω από τη γραφική της $\Delta(t)$. Επομένως, είναι δυνατόν να αποτυπώσουμε αυτή την περιοχή ως το άθροισμα των γεωμετρικών χώρων που στο Σχήμα 2.2, δηλώνονται ως Q_i , όπου $i=1,2,3, \dots, N(\tau)$, με το $N(\tau)$ να ορίζεται στη Σχέση (2.1). Επιπλέον, προσθέτουμε την τριγωνική περιοχή στο τέλος, που συμβολίζεται, όπως φαίνεται στο Σχήμα 2.2, με \bar{Q} ή έστω Q . Έτσι:

$$\Delta\tau = \frac{1}{c} (Q_1 + \dots + \sum_{i=2}^{N(\tau)} Q_i) = \frac{Q_1 + Q}{\tau} + \frac{N(\tau)-1}{\tau} \frac{1}{N(\tau)-1} \sum_{i=2}^{N(\tau)} Q_i \quad (2.7)$$

Η average age of information ισούται:

$$\Delta = \lim_{\tau \rightarrow \infty} \Delta\tau \quad (2.8)$$

Επίσης, λόγω του Σχήματος 2.2 και των σχέσεων (2.4),(2.5) μπορούμε να γράψουμε:

$$Q_i = \frac{1}{2} (T_i + Y_i)^2 - \frac{1}{2} T_i^2 = Y_i T_i + \frac{Y_i^2}{2} \quad (2.9)$$

Οπότε η μέση ηλικία (average age) προκύπτει:

$$\Delta = \frac{E[Q]}{E[Y]} = \frac{E[YT] + \frac{E[Y^2]}{2}}{E[Y]} \quad (2.10)$$

Αξίζει να παρατηρήσουμε πως έχουμε υποθέσει εργοδικότητα για τη διαδικασία $\Delta(t)$, αλλά όχι για τις μεταβλητές Y, T . Επειδή οι δύο τελευταίες είναι εξαρτημένες μεταβλητές, δεν είναι απλός ο υπολογισμός της average age, σε περίπτωση που δε γνωρίζουμε πως κατανέμονται.

Για μια δεδομένη τιμή ρυθμού εξυπηρέτησης, μειώνοντας το Y (interarrival times), αυξάνονται τα πακέτα που θα είναι κάθε στιγμή στο σύστημα με αποτέλεσμα να δημιουργείται μεγαλύτερος συνωστισμός πακέτων που οδηγεί σε μεγαλύτερο T (system time). Αντίθετα, για μεγάλα Y , η ουρά είναι πιο άδεια και έχουμε, άρα, μικρότερες καθυστερήσεις με μικρότερα T . Επομένως, βλέπουμε πως τα Y, T είναι αρνητικά συσχετισμένες μεταβλητές.

2.2.2 Peak Age και Average Peak Age

Υποθέτουμε Y_i ο χρόνος που μεσολαβεί μεταξύ δύο διαδοχικών αφίξεων πακέτων $i-1, i$ και T_i , να είναι ο χρόνος στο σύστημα για το i -οστό μεταδιδόμενο πακέτο. Η Peak Age of Information, P AoI, θα είναι:

$$A_i = Y_i + T_i \quad (2.11)$$

Από το Σχήμα 2.2 μπορούμε να δούμε ακριβώς πως προκύπτει η Peak Age of Information ως το άθροισμα των χρόνων Y_i, T_i . Η P AoI προκύπτει από μια στοχαστική διαδικασία σε διακριτό χρόνο και παίρνει τιμές από το σύνολο $L = \{t'_0, t'_1, t'_2, \dots, t'_{N(\tau)}\}$. Καταλαβαίνουμε πως είναι σημαντικό, όπως και στην περίπτωση της AoI, να έχουμε μικρές τιμές P AoI, δηλαδή μικρή μέση τιμή P AoI (Average Peak Age of Information) ώστε να διατηρούμε φρέσκιες πληροφορίες.

Αν πάρουμε ως διάστημα παρατήρησης το $(0, \tau)$ με $N(\tau)$ τον αριθμό των δειγμάτων που έχουν εξυπηρετηθεί μέχρι τη στιγμή τ , η time average peak Age of Information είναι:

$$A_\tau = \frac{1}{N(\tau)} \sum_{i=2}^{N(\tau)} A_i \quad (2.12)$$

Έτσι, η average peak age of information ισούται:

$$A = \lim_{\tau \rightarrow \infty} A_\tau \quad (2.13)$$

Χρησιμοποιώντας τη Σχέση (2.11), προκύπτει:

$$A = E[Y + T] \quad (2.14)$$

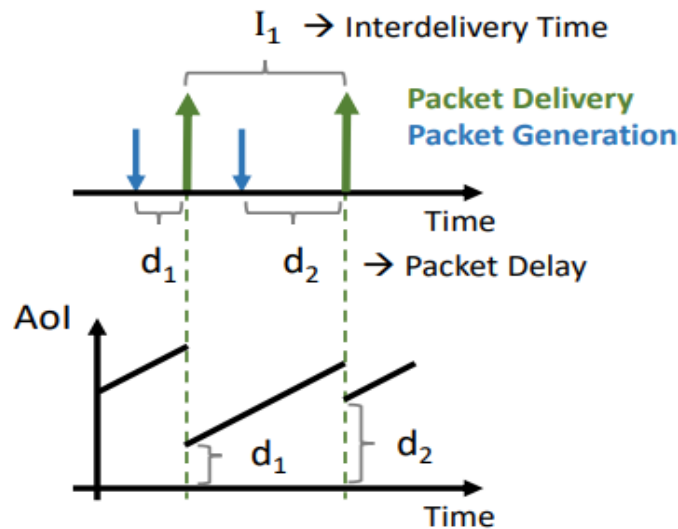
όπου $T = W + S$ με W : χρόνος αναμονής και S : χρόνος εξυπηρέτησης.

Να σημειωθεί ότι η peak age μπορεί να χρησιμοποιηθεί σε εφαρμογές που μας ενδιαφέρει η χειρότερη περίπτωση ηλικίας ή έχουμε περιορισμό για τη μέγιστη ηλικία.

2.3 Η AoI σε Σχέση με Άλλες Παραμέτρους

Δύο βασικές παράμετροι που επηρεάζουν την ηλικία της πληροφορίας είναι η καθυστέρηση (delay) και ο χρόνος μεταξύ διαδοχικών παραλαβών των πακέτων (interdelivery time). Συγκεκριμένα, η καθυστέρηση είναι ο χρόνος που μεσολαβεί από την δημιουργία μέχρι την παράδοση του πακέτου.

Στο Σχήμα 2.3, φαίνονται οι μετρικές που μας ενδιαφέρουν καθώς και πως συνδέονται με την AoI στο επίπεδο του χρόνου.



Σχήμα 2.3 – Χρονική απεικόνιση Age Of Information, Delay, Interdelivery Time

Ο έλεγχος μόνο ενός από τις δύο παραμέτρους, που σχολιάσαμε, δεν είναι αρκετός για την επίτευξη καλής AoI. Για παράδειγμα, σε ένα σύστημα ενός εξυπηρετητή που έχουμε υψηλό ρυθμό εξυπηρέτησης και χαμηλό ρυθμό αφίξεων, οδηγούμαστε σε μικρή καθυστέρηση, αφού η ουρά είναι συχνά άδεια. Ωστόσο, η ηλικία της πληροφορίας μπορεί να έχει μεγάλη τιμή, αφού οι αφίξεις γίνονται αραιότερα, με αποτέλεσμα η πληροφορία που φτάνει στον προορισμό να είναι ίσως ξεπερασμένη.

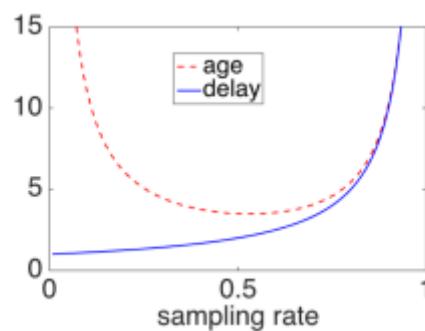
Στον παρακάτω πίνακα (Πίνακας 2.1), αποδεικνύεται αυτό που προαναφέρθηκε, με το αριθμητικό παράδειγμα μιας M/M/1 ουράς με ρυθμό εξυπηρέτησης $\mu=1$ όπου υψηλές τιμές ενός εκ των δύο από τα delay, interdelivery time οδηγούν σε υψηλή AoI παρ' ότι το άλλο μπορεί να έχει χαμηλή τιμή.

λ (pkt/sec)	E[delay] (sec)	E[interdelivery] (sec)	Average AoI (sec)
0.01	1.01	100.00	101.00
0.53	2.13	1.89	3.48
0.99	100.00	1.01	100.02

Πίνακας 2.1 – Μέση AoI για διάφορες τιμές των delay, interdelivery time σε ουρά M/M/1 με $\mu=1$ και μεταβλητό λ [14]

Βλέπουμε πως, η βέλτιστη τιμή για τη μέση ηλικία επιτυγχάνεται με ειδικό συνδυασμό των delay και interdelivery time, και όχι στις ελάχιστες ή μέγιστες τιμές τους. Αυτό συμβαίνει γιατί η AoI παίρνει καλές τιμές, όταν έχουμε μια σχετικά χαμηλή καθυστέρηση αλλά τακτική παράδοση πακέτων.

Αυτό αποτυπώνεται και στο παρακάτω σχήμα όπου παρατηρούμε την πορεία των τιμών των age, delay σε συνάρτηση με το ρυθμό παραγωγής δειγμάτων. Η ελάχιστη AoI προκύπτει για έναν «μεσαίο» ρυθμό δημιουργίας πακέτων και μικρή καθυστέρηση (όχι ελάχιστη).



Σχήμα 2.4 – Η ηλικία της πληροφορίας και η καθυστέρηση σε συνάρτηση με το ρυθμό δειγματοληψίας για M/M/1 FIFO ουρά

Κεφάλαιο 3: Μοντέλα AoI Συνεχούς Χρόνου

Στο Κεφάλαιο αυτό γίνεται μελέτη μοντέλων συνεχούς χρόνου, δηλαδή μοντέλων όπου ο χρόνος παραμονής σε μία κατάσταση ακολουθεί διαδικασία συνεχούς χρόνου. Αυτό σημαίνει πως η ηλικία της πληροφορίας στα μοντέλα αυτά ορίζεται ως συνεχής διαδικασία.

3.1 Μοντέλα με 1 Εξυπηρετητή

Αρχικά θα εξεταστούν ορισμένα απλά μοντέλα ενός εξυπηρετητή που ακολουθούν first-come-first-served (FCFS) πολιτική. Είναι τα M/M/1, M/D/1 και D/M/1 μοντέλα, τα οποία αναλύονται στο [1] και χρησιμοποιούνται ως συστήματα για την απλούστερη ανάλυση των ιδιοτήτων τους, καθώς και των μετρικών που μας ενδιαφέρουν.

3.1.1 M/M/1 - FIRST COME FIRST SERVED

Θα μελετήσουμε την περίπτωση ενός M/M/1 συστήματος τα πακέτα του οποίου ακολουθούν FCFS (First Come First Served) πολιτική. Το σύστημα περιέχει μία πηγή και έναν εξυπηρετητή με τη διαδικασία αφίξεων να είναι Poisson μέσο ρυθμό λ ενώ οι χρόνοι εξυπηρέτησης ακολουθούν εκθετική κατανομή με μέσο $1/\mu$. Υποθέτουμε ότι οι χρόνοι εξυπηρέτησης είναι ανεξάρτητοι μεταξύ τους και ανεξάρτητοι των interarrival times Y_i με $E[Y_i] = 1/\lambda$. Ταυτόχρονα, ο εξυπηρετητής ποτέ δεν μένει ανενεργός όταν υπάρχει πελάτης στο σύστημα (work conserving) ενώ έχει χρησιμοποίηση $\rho = \frac{\lambda}{\mu}$. Για τον υπολογισμό της ηλικίας της πληροφορίας (AoI) και βασιζόμενοι στη Σχέση (2.9), βλέπουμε πως πρέπει να υπολογιστούν πρώτα οι παράγοντες $E[YT]$ και $E[Y^2]$.

Δεδομένου ότι η Y είναι εκθετικά κατανομημένη μεταβλητή με μέσο ρυθμό λ , προκύπτει $E[Y^2] = 2/\lambda^2$. Για τον παράγοντα $E[YT]$ θα πρέπει να βρούμε το χρόνο του συστήματος, ο οποίος για μία ενημέρωση κατάστασης i είναι ίσος με:

$$T_i = W_i + S_i \quad (3.1)$$

όπου W_i είναι ο χρόνος αναμονής και S_i ο χρόνος εξυπηρέτησης ($=1/\mu$).

Εφόσον S_i, Y_i ανεξάρτητες μεταβλητές, προκύπτει:

$$E[T_i Y_i] = E[(W_i + S_i)Y_i] = E[W_i Y_i] + E[S_i]E[Y_i] \quad (3.2)$$

Μπορούμε να θεωρήσουμε ότι $W_i = (T_{i-1} - X_i)^+$ αφού ισχύει $W_i = 0$ αν το πακέτο $i-1$ έχει εξυπηρετηθεί όταν το πακέτο i παράγεται ενώ σε άλλη περίπτωση ισχύει:

$$W_i = T_{i-1} - X_i.$$

Γνωρίζοντας τη συνάρτηση πυκνότητας πιθανότητας:

$$f_T(t) = \mu(1 - \rho)e^{-\mu(1 - \rho)t}, \quad t \geq 0 \quad (3.3)$$

μπορούμε να υπολογίσουμε τον παράγοντα:

$$\begin{aligned} E[W_i | Y_i = y] &= \int_y^{\infty} (t - y) f_T(t) dt \\ &= \frac{e^{-\mu(1-\rho)y}}{\mu(1-\rho)} \end{aligned} \quad (3.4)$$

και έπειτα το

$$E[W_i T_i] = \int_y^{\infty} y E[W_i | Y_i = y] f_{Y_i}(y) dy = \frac{\rho}{\mu^2(1 - \rho)} \quad (3.5)$$

Έτσι από τις σχέσεις (2.10), (3.1), (3.5) προκύπτει ότι η μέση ηλικία της πληροφορίας είναι:

$$\Delta_{M/M/1} = \frac{1}{\mu} \left(1 + \frac{1}{\rho} + \frac{\rho^2}{1 - \rho} \right) \quad (3.6)$$

Αυτό που επιθυμούμε είναι να ελαχιστοποιηθεί η μέση ηλικία της πληροφορίας χωρίς να επηρεαστεί η χρησιμοποίηση ρ του εξυπηρετητή. Θεωρώντας δεδομένες τιμές για τα μ, λ και υποθέτοντας ότι εξετάζουμε τυχαίες στιγμές της διαδικασίας $X(t)$, η ζητούμενη τιμή για τη χρησιμοποίηση είναι $\rho^* \approx 0.53$ όπως εξάγεται από τη Σχέση (3.6).

Βλέπουμε, λοιπόν, πως η ελάχιστη μέση ΑοΙ δεν επιτυγχάνεται ούτε έχοντας μέγιστη απόδοση (όταν η χρησιμοποίηση τείνει στο 1) ούτε έχοντας ελάχιστη καθυστέρηση (όταν η χρησιμοποίηση τείνει στο 0).

3.1.2 M/D/1 - FIRST COME FIRST SERVED

Στο M/D/1 σύστημα έχουμε ενημερώσεις κατάστασης που παράγονται βάσει της διαδικασίας Poisson με μέσο ρυθμό λ και ντετερμινιστικό χρόνο εξυπηρέτησης. Αυτό σημαίνει πως ισχύει $S_i = D$ για κάθε πακέτο i όπου D σταθερή τιμή. Για να προσδιοριστεί, εκ νέου, η μέση ηλικία, χρειαζόμαστε τον παράγοντα $E[YT]$ δεδομένου ότι $E[Y^2] = 2/\lambda^2$ αφού η Y είναι εκθετικά κατανομημένη μεταβλητή με μέσο ρυθμό λ .

Θεωρώντας ότι ο χρόνος του συστήματος είναι:

$$T_i = W_i + D \quad (3.7)$$

Προκύπτει:

$$E[T_i Y_i] = E[W_i Y_i] + D E[Y_i] \quad (3.8)$$

με $E[Y_i] = 1/\lambda$.

Μπορούμε να γράψουμε:

$$E[W_i | Y_i = y] = E[(T - y)^+] = E[(W + D - y)^+] \quad (3.9)$$

όπου γνωρίζοντας πως $E[W] = \frac{E[S]\rho}{2(1-\rho)}$, μπορεί να υπολογιστεί ο παράγοντας $E[W_i | Y_i = y]$ και ως αποτέλεσμα να εξάγουμε το $E[W_i Y_i]$.

Λόγω της περιπλοκότητας της ανάλυσης δεν παρέχεται η ακριβής έκφραση της

ηλικίας της πληροφορίας για ένα σύστημα M/D/1 [1].

Η χρησιμοποίηση ρ που ελαχιστοποιεί τη μέση AoI είναι ίση με $\rho^* \approx 0.625$ όπως βρίσκεται στο [1].

3.1.3 D/M/1 - FIRST COME FIRST SERVED

Στο D/M/1 σύστημα έχουμε ενημερώσεις κατάστασης που παράγονται σε μία ντετερμινιστική περίοδο D ενώ οι χρόνοι εξυπηρέτησης είναι εκθετικά κατανεμημένοι με μέση τιμή $1/\mu$.

Ισχύει: $E[Y] = D$, $E[Y^2]/2 = D^2/2$, $E[YT] = DE[T]$

Οπότε αρκεί να υπολογιστεί η αναμενόμενη τιμή $E[T]$ που αφορά το χρόνο συστήματος T. Θεωρώντας ότι ο χρόνος του συστήματος είναι:

$$T_i = W_i + S \quad (3.10)$$

ο μέσος χρόνος του συστήματος μπορεί να γραφεί:

$$E[T] = E[W] + E[S] = \frac{1}{\mu} + \frac{\beta}{\mu(1-\beta)} \quad (3.11)$$

όπου $0 \leq \beta \leq 1$ είναι η λύση της εξίσωσης $\beta = L_x(\mu(1-\beta))$

με L_x να είναι ο μετασχηματισμός Laplace της κατανομής του χρόνου μεταξύ δύο συνεχόμενων επιτυχημένων αφίξεων [4].

Έτσι, προκύπτει ότι η μέση ηλικία της πληροφορίας είναι:

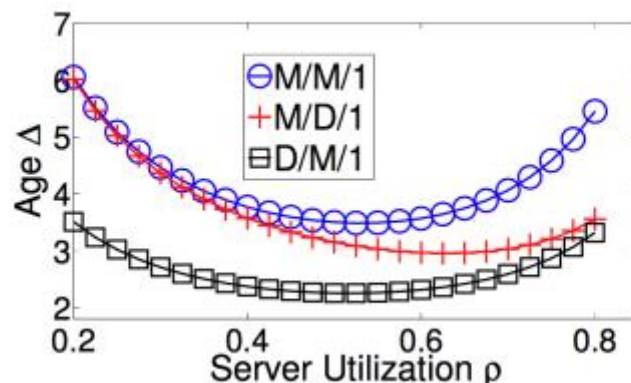
$$\Delta_{M/M/1} = \frac{1}{D} + \left(\frac{D^2}{2} + DE[T] \right) = \frac{1}{\mu} \left(\frac{1}{2\rho} + \frac{1}{1-\beta} \right) \quad (3.12)$$

Για την ελαχιστοποίηση της μέσης AoI, η χρησιμοποίηση είναι ίση με $\rho^* \approx 0.515$.

3.1.4 Σύγκριση M/M/1, M/D/1, D/M/1

Συνοπτικά τα 3 διαφορετικά μοντέλα που εξετάστηκαν, συγκρίνονται ως προς τη μέση AoI σε συνάρτηση με τη χρησιμοποίηση για σταθερό ρυθμό εξυπηρέτησης ίσο

με $\mu=1$.



Σχήμα 3.1- Η μέση ηλικία της πληροφορίας για τα M/M/1, M/D/1, D/M/1 ως προς τη χρησιμοποίηση για $\mu=1$ [1]

Βλέπουμε πως το μοντέλο D/M/1 επιτυγχάνει αρκετά χαμηλή μέση AoI σε σχέση με τα άλλα 2 μοντέλα για όλες τις τιμές ρ . Όσο αυξάνεται η τιμή του ρ , παρατηρούμε πως οι τιμές του M/D/1 πλησιάζουν εκείνες του D/M/1 ενώ απομακρύνονται από αυτές του M/M/1 που αρχικά ήταν σχεδόν ίδιες. Αυτό συμβαίνει διότι για μεγαλύτερη χρησιμοποίηση, έχουμε περισσότερα πακέτα σε αναμονή, κατάσταση στην οποία συμπεριφέρεται καλύτερα το M/D/1 από το M/M/1, αφού ο ντετερμινισμός στην εξυπηρέτηση στο M/D/1 μειώνει την ουρά.

3.2 Μοντέλα με Περισσότερους Εξυπηρετητές

Μέχρι στιγμής έχουμε εξετάσει μοντέλα τα οποία έχουν έναν εξυπηρετητή, όπου η λήψη του πακέτου από τον προορισμό γίνεται με σειρά. Ωστόσο, τα δίκτυα μπορεί να είναι δυναμικά και οι τοπολογίες τους να αλλάζουν, αφού οι συνδέσεις μεταξύ των κόμβων μπορεί να είναι ρυθμιζόμενες. Έτσι, η λήψη των πακέτων μπορεί να μην συμβαίνει με τη σειρά που γίνεται η εκπομπή τους, με αποτέλεσμα κάποια πακέτα να περιέχουν παλαιότερη πληροφορία από αυτή που έχει ληφθεί ήδη από άλλα.

Αυτό δημιουργεί την ανάγκη να εξεταστούν μοντέλα με περισσότερους από έναν εξυπηρετητές, όπως είναι τα M/M/c με $c>1$, για την αντιμετώπιση διάφορων φαινομένων σαν των πολλαπλών μονοπατιών στη μετάδοση πακέτων.

3.2.1 Informative – Obsolete Πακέτα

Πρώτα, καλό είναι να οριστεί ποια πακέτα ονομάζουμε πληροφοριακά (informative). Θεωρούμε ένα σύστημα που περιέχει ένα σύνδεσμο επικοινωνίας πηγής-προορισμού, που χωρίζονται από ένα δικτυακό νέφος. Ορίζουμε το informative πακέτο ως ένα πακέτο m που “κουβαλάει” πιο νέα πληροφορία σε σύγκριση με τα πακέτα που έχουν φτάσει νωρίτερα στον προορισμό. Αν θεωρήσουμε Y_i ως interarrival times και S_i ως service times του update, τότε για να είναι ένα πακέτο πληροφοριακό πρέπει:

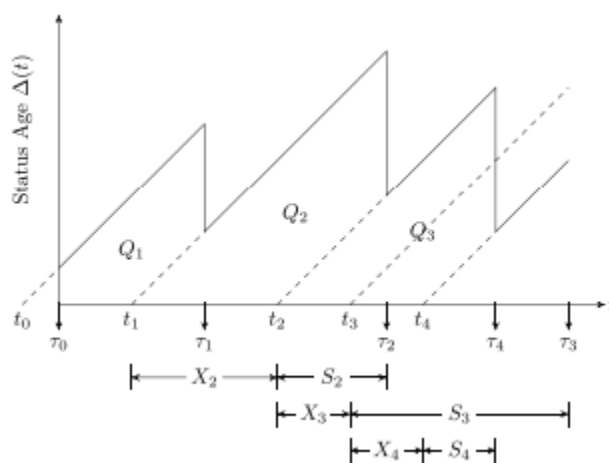
$$S_m < S_r + \sum_{a=m+1}^r Y_a, \forall r > m \quad (3.13)$$

Αντίθετα, ένα πακέτο m θεωρείται ξεπερασμένο (obsolete) αν υπάρχει πακέτο k που δημιουργήθηκε μετά από το m ($t_m < t_k$), αλλά έφτασε στον προορισμό νωρίτερα ($t_k < t_m$).

Ένα informative πακέτο, δεν καθίσταται obsolete.

3.2.2 M/M/∞

Αρχικά, θα δούμε το M/M/∞ μοντέλο, όπου έχουμε μια πηγή και άπειρους εξυπηρετητές με τη διαδικασία αφίξεων να είναι Poisson μέσου ρυθμού λ και τους χρόνους εξυπηρέτησης να ακολουθούν εκθετική κατανομή με μέσο $1/\mu$. Κατά τη μετάδοση, όταν η πηγή παράγει ένα πακέτο, εκείνο περιέχει τωρινή πληροφορία και ξεκινά αμέσως η εξυπηρέτησή του, μιας και το σύστημα έχει άπειρους εξυπηρετητές.



Σχήμα 3.2 - Η ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το μοντέλο M/M/∞ [8]

Λόγω της πολυπλοκότητας του υπολογισμού της average AoI για το μοντέλο $M/M/\infty$, χρησιμοποιούμε όρια για να την προσδιορίσουμε προσεγγιστικά.

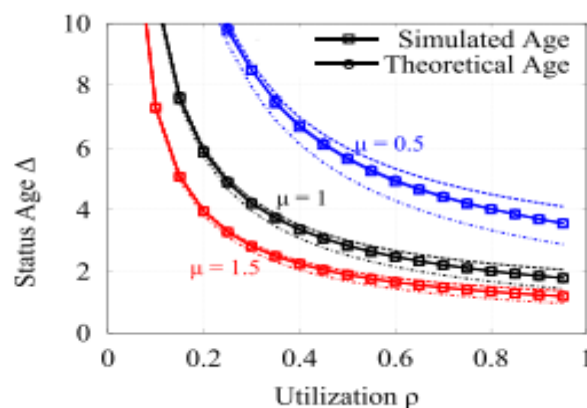
Για το ανώτερο όριο (upper bound), σύμφωνα με το [8], υπολογίζουμε τον μέσο για τις τραπεζοειδής περιοχές όλων των πακέτων, είτε είναι informative είτε όχι. Προκύπτει:

$$\Delta_{U,M/M/\infty} = \frac{1}{\lambda} + \frac{1}{\mu} \quad (3.14)$$

ενώ για το κατώτερο όριο (lower bound), θεωρούμε χρόνο εξυπηρέτησης ενός πακέτου μικρότερο από το interarrival χρόνο του επόμενου πακέτου, δηλαδή $S_i < Y_{i+1}$. Αυτό σημαίνει πως το νέο update δημιουργήθηκε αφού τελείωσε η εξυπηρέτηση του προηγούμενου. Έτσι προκύπτει [8]:

$$\Delta_{L,M/M/\infty} = \frac{1}{\lambda} + \frac{1}{\lambda + \mu} - \frac{\lambda \mu}{(\lambda + \mu)^3} \quad (3.15)$$

Να σημειωθεί ότι τα όρια γίνονται πιο “στενά”, δηλαδή έχουν μικρότερη απόκλιση μεταξύ τους, όταν η χρησιμοποίηση ρ είναι μικρότερη, ενώ καθώς εκείνη αυξάνεται ($\rho \uparrow$) και μικραίνει ο ρυθμός εξυπηρέτησης μ τα όρια αποκλίνουν περισσότερο.



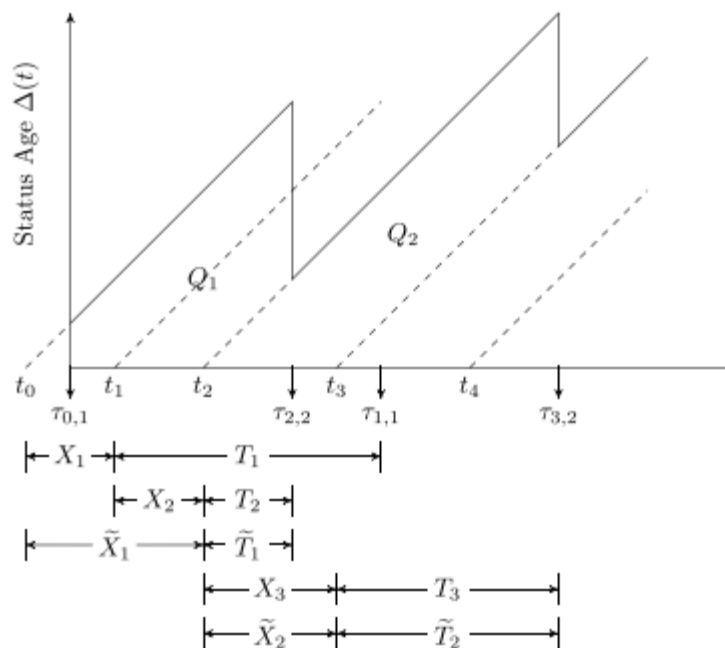
Σχήμα 3.3 - Η ηλικία της πληροφορίας σε συνάρτηση με τη χρησιμοποίηση για διάφορες τιμές του μ για το μοντέλο $M/M/\infty$ [8]

Στο Σχήμα 3.3 βλέπουμε πως καθώς η χρησιμοποίηση αυξάνεται έχουμε και μείωση

της ηλικίας της πληροφορίας, αφού λόγω των άπειρων εξυπηρετητών, τα πακέτα εξυπηρετούνται άμεσα. Έτσι, το αυξημένο πλήθος τους οδηγεί σε συνεχή status updates στον προορισμό, δηλαδή πιο πρόσφατη πληροφορία, άρα και μικρότερη μέση AoI.

3.2.3 M/M/2

Εξετάζοντας το μοντέλο M/M/2, υποθέτουμε εκ νέου εκθετικούς interarrival χρόνους Y_i με μέσο $1/\lambda$ και εκθετικούς χρόνους εξυπηρέτησης S_i με μέσο $1/\mu$, αλλά αυτή τη φορά έχουμε 2 εξυπηρετητές. Παρ' όλο που έχουμε είσοδο στην ουρά με τη σειρά, δεν είναι βέβαιο σε ποιον εξυπηρετητή θα εισέλθει κάθε πακέτο άρα και ποιο πακέτο θα τελειώσει πρώτα την εξυπηρέτησή του. Έτσι, γίνεται και ο υπολογισμός της AoI του μοντέλου M/M/2 περίπλοκος, καταφεύγοντας ξανά στη λύση των upper και lower ορίων για το προσδιορισμό της.



Σχήμα 3.4 - Η ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το μοντέλο M/M/2 [8]

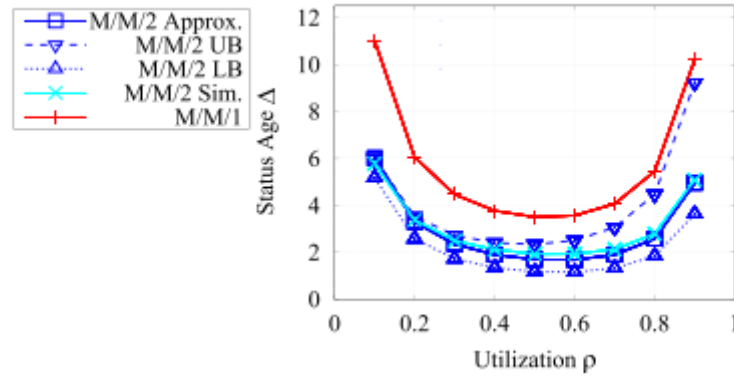
Με όμοια διαδικασία του μοντέλου M/M/∞, αλλά αυτή τη φορά για το M/M/2 υπολογίζεται το upper bound ίσο με:

$$\Delta_{U,M/M/2} = \frac{1}{\mu} \left(1 + \frac{1}{2\rho} + \frac{2\mu\rho^3}{(1+\rho)(1-\rho)} \right) \quad (3.16)$$

ενώ για το lower bound προκύπτει πιο περίπλοκος τύπος εξαρτώμενος από το χρόνο \tilde{Y}_a ενός informative πακέτου, και το $\tilde{\lambda} = \lambda (p_a + p_b)$. Ορίζονται ως p_a η πιθανότητα ενός informative πακέτου να έχει το προηγούμενο του informative, ενώ ως p_b η πιθανότητα ενός informative πακέτου να καθιστά το προηγούμενο του obsolete. Σύμφωνα με το [8], ισούται με:

$$\Delta_{L,M/M/2} = \tilde{\lambda} (E[\tilde{W} \tilde{Y}_a] + E[\tilde{Y}_a]E[\tilde{S}] + \frac{E[\tilde{Y}_a^2]}{2}) \quad (3.17)$$

Στο παρακάτω σχήμα (Σχήμα 3.5) βλέπουμε τη μείωση της ηλικίας της πληροφορίας για το μοντέλο M/M/2 σε σύγκριση με το M/M/1.



Σχήμα 3.5 - Η ηλικία της πληροφορίας σε συνάρτηση με τη χρησιμοποίηση για τα μοντέλα M/M/2, M/M/1 για $\mu=1$ [8]

Παρατηρούμε πως οι τιμές της average AoI για το M/M/2 είναι σχεδόν στο μισό σε σχέση με το M/M/1. Αυτό συμβαίνει λόγω του ενός επιπλέον εξυπηρετητή στο M/M/2 που αποσυμφορίζει την κίνηση στην ουρά των πακέτων, με αποτέλεσμα την πιο άμεση εξυπηρέτηση, άρα και την απόκτηση πιο “φρέσκιας” πληροφορίας από τον προορισμό. Βέβαια, είναι δυνατόν, να μεταφερθούν obsolete πακέτα, που δε χρησιμεύουν σε τίποτα για την AoI αλλά καταναλώνουν πόρους του δικτύου.

Κεφάλαιο 4: Πολιτικές και Τεχνικές Μείωσης ΑοΙ

4.1 LCFS Πολιτική

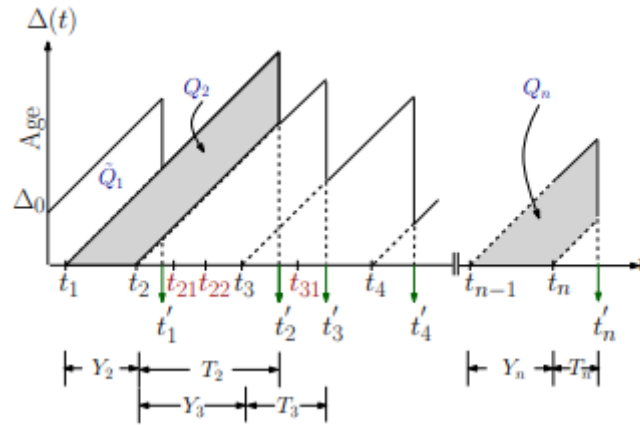
Στο κεφάλαιο αυτό θα εξετάσουμε την περίπτωση όπου έχουμε LCFS (Last Come First Served) πολιτική. Παραπάνω είδαμε πως στη FCFS πολιτική, στέλνοντας όσο το δυνατόν γρηγορότερα τα πακέτα, μπορεί η χρησιμοποίηση να μεγιστοποιείται, ωστόσο είναι πιθανόν να οδηγηθούμε σε μεγάλες τιμές ΑοΙ. Αυτό εξηγείται από το γεγονός ότι, τα πακέτα μένουν σε αναμονή μέχρι να εξυπηρετηθούν λόγω της συμφόρησης που παρουσιάζεται. Μια ενδεχόμενη λύση σε αυτή την κατάσταση θα ήταν η μείωση του ρυθμού παραγωγής των updates, κάτι που όμως ίσως οδηγήσει σε χρονικά ξεπερασμένες ενημερώσεις. Έτσι, δημιουργείται η ανάγκη να ακολουθήσουμε μια άλλη πολιτική, κατά την οποία η εξυπηρέτηση των νεότερων πακέτων θα είναι άμεση. Η πολιτική LCFS δίνει αυτή τη δυνατότητα, δεδομένου ότι η κατάσταση είναι Μαρκοβιανή, και ως εκ τούτου δεν υπάρχει όφελος από τη λήψη παλαιότερων πακέτων [5].

Θα μελετηθούν δύο διαφορετικές περιπτώσεις για την LCFS, με δικαίωμα προτίμησης (with preemption) και χωρίς δικαίωμα προτίμησης (without preemption).

4.1.1 LCFS Without Preemption

Όταν δεν υπάρχει δικαίωμα προτίμησης, το νέο πακέτο μπορεί να αντικαταστήσει κάθε παλαιότερο πακέτο που περιμένει στην ουρά, αλλά πρέπει να περιμένει να τελειώσει η πιθανή τρέχουσα εξυπηρέτηση ενός πακέτου.

Παρακάτω παρουσιάζεται η μέση ΑοΙ καθώς μια πηγή στέλνει ενημερώσεις σε έναν monitor για LCFS without preemption. Το σχήμα 4.1 που ακολουθεί δείχνει την τιμή της μέσης ηλικίας της πληροφορίας σε σχέση με το χρόνο.



Σχήμα 4.1 – Παράδειγμα πορείας average Aol σε monitor με LCFS without preemption [5]

Βλέπουμε πως με t_i όπου $i=1,2,3,\dots,n$ συμβολίζονται οι χρόνοι που δημιουργούνται τα αντίστοιχα πακέτα i ενώ όμοια με t'_i οι αντίστοιχοι χρόνοι εξυπηρέτησης τους. Ανάμεσα στους χρόνους t'_i και t'_{i+1} μπορεί να δημιουργούνται πακέτα από άλλες πηγές που ίσως εξυπηρετηθούν αλλά και πακέτα από την πηγή που εξετάζουμε, τα οποία όμως δεν πρόλαβαν να εξυπηρετηθούν, αφού αντικαταστάθηκαν από νεότερα.

Όπως είναι λογικό, η μέση ηλικία αυξάνεται γραμμικά με το χρόνο για όσο δεν έχουμε εξυπηρέτηση πακέτου που επηρεάζει την ηλικία. Προφανώς, τα πακέτα αυτά που επηρεάζουν την ηλικία είναι εκείνα από την πηγή που μας ενδιαφέρει και ολοκληρώνουν την εξυπηρέτηση τους. Μάλιστα εκείνη τη στιγμή t'_i , η τιμή $\Delta(t'_i)$ πέφτει και γίνεται ίση με $T_i = t'_i - t_i$, που είναι και ο χρόνος που το πακέτο i βρίσκεται στο σύστημα.

Θεωρούμε $Y_i = t_i - t_{i-1}$ το χρόνο μεταξύ της δημιουργίας δύο τέτοιων πακέτων, οπότε από το Σχήμα 4.1 προκύπτει:

$$Q_i = \frac{(Y_i + T_i)^2}{2} - \frac{T_i^2}{2} \quad (4.1)$$

Παίρνουμε ως διάστημα παρατήρησης το $(0, \tau)$ και η time average age είναι:

$$\Delta\tau = \frac{1}{\tau} \int_0^\tau \Delta(t) dt \quad (4.2)$$

Η μέση ηλικία που προκύπτει, όπως αναφέρεται στο [5], είναι:

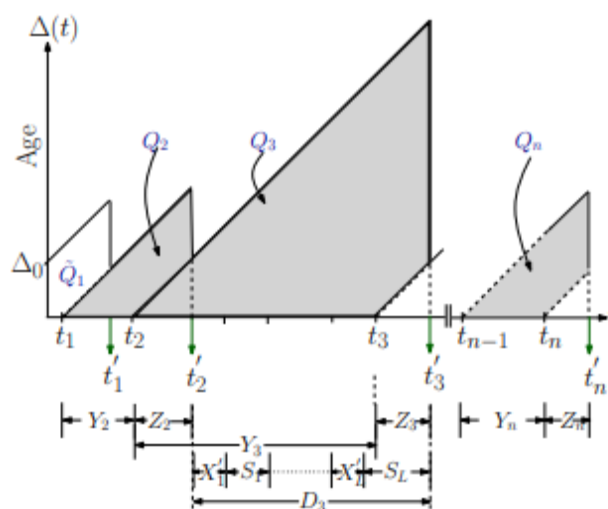
$$\Delta = \lim_{\tau \rightarrow \infty} \Delta\tau = \frac{1}{E[Y]} \left[\frac{E[Y^2]}{2} + E[YT] \right] \quad (4.3)$$

όπου $E[\cdot]$ ο τελεστής αναμενόμενης τιμής.

4.1.2 LCFS With Preemption

Στην περίπτωση της πολιτικής LCFS with preemption, το νέο πακέτο μετά τη δημιουργία του εξυπηρετείται άμεσα αφού αντικαθιστά κάθε πακέτο που υπάρχει είτε στην ουρά είτε εξυπηρετείται εκείνη τη στιγμή. Τα πακέτα που έρχονται ή αντικαθίστανται δεν είναι ανάγκη να είναι από την ίδια πηγή.

Όμοια με πριν, παρακάτω (Σχήμα 4.2) παρουσιάζεται η μέση AoI, όπως εκείνη προκύπτει για μία πηγή ενδιαφέροντος η οποία στέλνει updates σε έναν monitor αλλά αυτή τη φορά με πολιτική LCFS with preemption.



Σχήμα 4.2 – Παράδειγμα πορείας average AoI με LCFS with preemption [5]

Για τον υπολογισμό της μέσης ηλικίας της πληροφορίας, κρατάμε τους ίδιους υπάρχοντες συμβολισμούς όπως στο Σχήμα 4.1, αλλά βλέπουμε κι άλλες νέες

μεταβλητές. Συμβολίζουμε με D_i το χρονικό διάστημα μεταξύ των αφίξεων των πακέτων $i-1, i$. Δηλαδή, $D_i = t'_i - t'_{i-1}$.

Κάθε χρονικό διάστημα D_i περιέχει ορισμένα blocks του εξυπηρετητή που να είναι άπραγος και ακολουθείται από διάστημα του εξυπηρετητή που να είναι απασχολημένος. Αν το σύστημα περιέχει μία πηγή, τότε το D_i περιέχει ένα block. Εδώ βλέπουμε πως το D_3 περιέχει L blocks με το διάστημα απραγίας του εξυπηρετητή στο k -οστό block να ορίζεται ως X'_k και το αντίστοιχο διάστημα απασχόλησης του με S_k . Έτσι το συνολικό μήκος κάθε block είναι ίσο με:

$$B_k = X'_k + S_k$$

το σύνολο των οποίων συνθέτει το διάστημα: $D_i = \sum_{k=1}^L D_i = \sum_{k=1}^L (X'_k + S_k)$.

Η μέση ηλικία της πληροφορίας [5] που προκύπτει είναι:

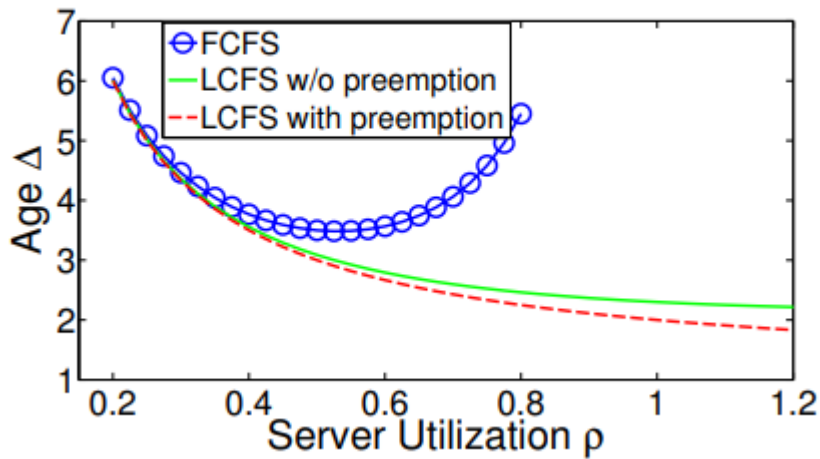
$$\Delta = \frac{1}{E[Y]} \left[\frac{E[Y^2]}{2} + E[YZ] \right] \quad (4.4)$$

ή

$$\Delta = E[Z] + \frac{E[D^2]}{2E[D]} + \frac{Var[Z] - Cov[DZ]}{E[D]} \quad (4.5)$$

4.1.3 Σύγκριση FCFS, LCFS

Παρατηρώντας το Σχήμα 4.3, είναι φανερό πως η LCFS with preemption επιτυγχάνει χαμηλότερες τιμές για την μέση AoI, σε σχέση με τις άλλες πολιτικές όσο η χρησιμοποίηση του εξυπηρετητή αυξάνεται. Αυτό είναι λογικό, αφού γίνεται αντικατάσταση των παλαιότερων πακέτων από νεότερα τόσο στην ουρά όσο και κατά την εξυπηρέτηση, κάτι πολύ ωφέλιμο όταν υπάρχουν πολλά πακέτα στο σύστημα.



Σχήμα 4.3 – Η ηλικία της πληροφορίας (σε έναν monitor) σε σχέση με τη χρησιμοποίηση του εξυπηρετητή για ένα μοντέλο $M/M/1$ FCFS, LCFS με και χωρίς δικαίωμα προτίμησης για $\mu=1$. [5]

Αντίθετα, στη FCFS κάθε πακέτο περιμένει την εξυπηρέτηση του σεβόμενο τη σειρά που καταφθάνει, με αποτέλεσμα όταν είναι πολλά πακέτα στην ουρά ($\rho \rightarrow 1$), να καθυστερεί να γίνει. Έτσι, η πληροφορία που θα προσφέρει το πακέτο στον προορισμό όταν εξυπηρετηθεί θα είναι ξεπερασμένη με αποτέλεσμα να έχουμε υψηλή μέση AoI. Ωστόσο, έως μία τιμή όπου ο ρυθμός εξυπηρέτησης (μ) είναι διπλάσιος από το ρυθμό αφίξεων (λ), βλέπουμε μείωση της ηλικίας, αφού τα πακέτα εξυπηρετούνται γρηγορότερα, κάτι που δε συμβαίνει καθώς το λ τείνει στο μ .

Βλέπουμε, επίσης, πως παρ' ότι δεν υπάρχουν μεγάλες διαφορές ανάμεσα στις δύο διαφορετικές πολιτικές LCFS για μικρές τιμές της χρησιμοποίησης, η απόκλιση τους, όσον αφορά την ηλικία, μεγαλώνει με την αύξηση της. Μάλιστα, σύμφωνα με το [5], όταν $\lambda \rightarrow \infty$ (άρα και $\rho \rightarrow \infty$), η LCFS with preemption επιτυγχάνει τη μισή ηλικία από εκείνη without preemption. Προφανώς, για το γεγονός αυτό ευθύνεται η επιπλέον “δυνατότητα” που δίνεται στη LCFS with preemption να αντικαθίστανται τα πακέτα που βρίσκονται υπό εξυπηρέτηση και όχι μόνο εκείνα στην ουρά όπως στη LCFS without preemption.

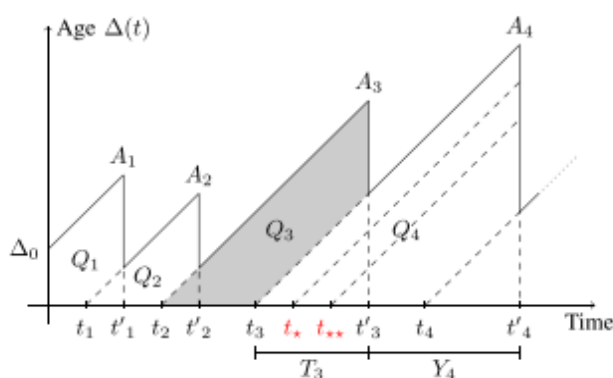
4.2 Διαχείριση Πακέτων

Όπως είδαμε και προηγουμένως, βασική προϋπόθεση για να υπάρχει χαμηλή ηλικία της πληροφορίας είναι να αποφύγουμε τη συμφόρηση των πακέτων και να κρατηθούν εκείνα που είναι πληροφοριακά (informative packets). Για να ελεγχθεί, λοιπόν, η κίνηση των πακέτων μέσα στο δίκτυο μπορεί να χρησιμοποιηθεί η διαχείριση των πακέτων (packet management) ως μια τεχνική απόρριψης ή αντικατάστασης εκείνων, ώστε να υπάρχει αποτελεσματικότερη χρησιμοποίηση των πόρων.

4.2.1 Ανάλυση Διαχείρισης Πακέτων

Για αρχή, θα γίνει η ανάλυση 3 διαφορετικών πολιτικών συγκρίνοντας τα αποτελέσματά τους [2]. Θεωρούμε αρχή FCFS, αφίξεις Poisson και εκθετικά κατανομημένους χρόνους εξυπηρέτησης για όλες τις παρακάτω πολιτικές:

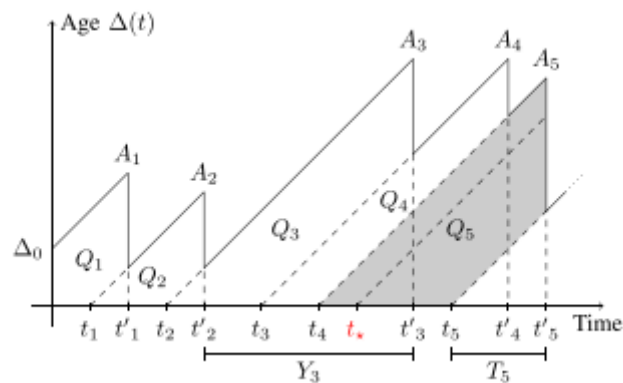
1) Για την 1^η πολιτική, αν ένα πακέτο φτάσει ενώ εξυπηρετείται κάποιο άλλο, τότε το νέο πακέτο απορρίπτεται. Τα πακέτα, όμως, που βρίσκουν τον εξυπηρετητή άδειο, εξυπηρετούνται κατευθείαν. Το μοντέλο αυτό συμβολίζεται ως M/M/1/1 αφού έχουμε 1 εξυπηρετητή και 1 πελάτη χωρητικότητα.



Σχήμα 4.4 - Η μέση ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το μοντέλο M/M/1/1 [2]

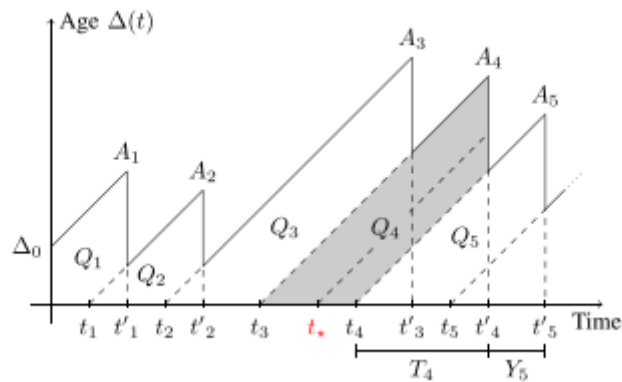
2) Για την 2^η πολιτική, αν ένα πακέτο φτάσει ενώ εξυπηρετείται κάποιο άλλο, αλλά η ουρά είναι άδεια τότε μπορεί να κρατηθεί σε αυτήν περιμένοντας την εξυπηρέτησή του. Ωστόσο, αν και στη μοναδική θέση της ουράς βρίσκεται άλλο πακέτο, τότε το νέο απορρίπτεται. Αντίθετα, τα πακέτα που βρίσκουν τον εξυπηρετητή άδειο, εξυπηρετούνται άμεσα. Το μοντέλο αυτό συμβολίζεται ως M/M/1/2 αφού έχουμε 1

εξυπηρετητή και 2 πελάτες χωρητικότητα.



Σχήμα 4.5 - Η μέση ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το μοντέλο M/M/1/2 [2]

3) Για την 3^η πολιτική, σε περίπτωση που φτάσει νεότερο πακέτο από αυτό που περιμένει στην ουρά να εξυπηρετηθεί, τότε το αντικαθιστά. Βέβαια, δεν αντικαθιστά και εκείνο που εξυπηρετείται εκείνη την ώρα, ενώ τα πακέτα που βρίσκουν τον εξυπηρετητή άδειο, εξυπηρετούνται άμεσα. Το μοντέλο αυτό συμβολίζεται ως M/M/1/2*.



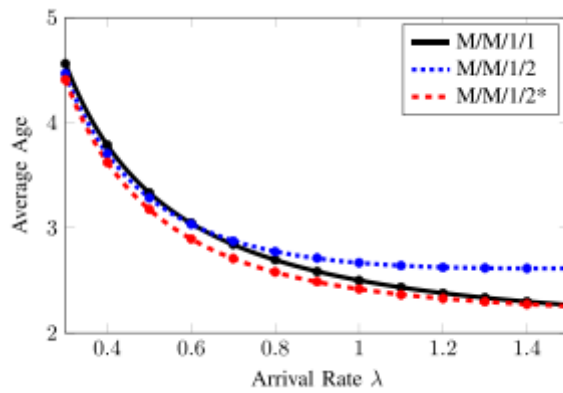
Σχήμα 4.6 - Η μέση ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το μοντέλο M/M/1/2* [2]

Η μέση ηλικία της πληροφορίας για καθένα από τα μοντέλα, όπως προκύπτει από το [2], είναι:

$$\Delta_{M/M/1/1} = \frac{1}{\lambda} + \frac{2}{\mu} - \frac{1}{\lambda + \mu} \quad (4.6)$$

$$\Delta_{M/M/1/2} = \frac{1}{\lambda} + \frac{3}{\mu} - \frac{2(\lambda + \mu)}{\lambda^2 + \lambda\mu + \mu^2} \quad (4.7)$$

$$\Delta_{M/M/1/2^*} = \frac{1}{\lambda} + \frac{2}{\mu} + \frac{\lambda}{(\lambda + \mu)^2} + \frac{1}{\lambda + \mu} - \frac{2(\lambda + \mu)}{\lambda^2 + \lambda\mu + \mu^2} \quad (4.8)$$



Σχήμα 4.7 - Η μέση ηλικία της πληροφορίας σε συνάρτηση με το ρυθμό αφίξεων λ για τα μοντέλα M/M/1/1, M/M/1/2, M/M/1/2* για σταθερό μ=1 [2]

Παρατηρώντας, στο Σχήμα 4.7, την average AoI ως προς το ρυθμό αφίξεων για μ=1, βλέπουμε πως καθώς αυξάνεται το λ, σημειώνεται μικρότερη average age για τα μοντέλα M/M/1/1 και M/M/1/2* σε σχέση με το M/M/1/2. Ελάχιστες τιμές, για πεπερασμένα λ, έχουμε με την 3η πολιτική (M/M/1/2*) όπου γίνεται αντικατάσταση των παλαιότερων πακέτων στην ουρά.

Αξίζει να αναφερθούμε και στο τι συμβαίνει στην περίπτωση όπου το λ απειρίζεται, αφού ταυτίζονται οι τιμές της average AoI για τα M/M/1/1 και M/M/1/2*.

Πιο συγκεκριμένα:

$$\lim_{\lambda \rightarrow \infty} \Delta_{M/M/1/1} = \lim_{\lambda \rightarrow \infty} \Delta_{M/M/1/2^*} = \frac{2}{\mu} \quad (4.9)$$

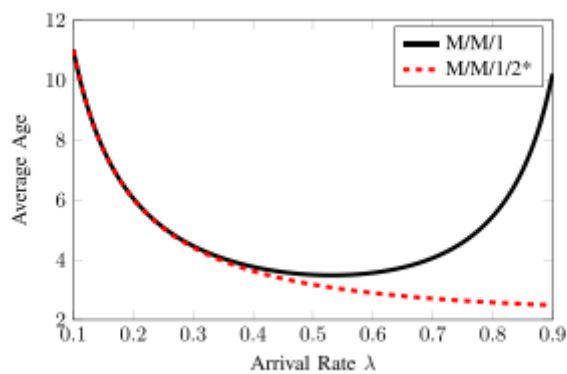
ενώ

$$\lim_{\lambda \rightarrow \infty} \Delta_{M/M/1/2} = \frac{3}{\mu} \quad (4.10)$$

Είναι εμφανές πως η average age για το M/M/1/2, όταν $\lambda \rightarrow \infty$, είναι μεγαλύτερη σε σχέση με τα υπόλοιπα δύο μοντέλα. Για πολύ μεγάλες τιμές του λ , έχουμε συνεχείς αφίξεις νέων πακέτων με αποτέλεσμα κάθε στιγμή να έχουμε διαθέσιμα «φρέσκα» πακέτα. Αυτό σημαίνει πως δε συμφέρει να έχουμε πακέτο, που δεν αντικαθίσταται, να περιμένει στην ουρά, επειδή δεν θα είναι τόσο πρόσφατη η πληροφορία του όταν εξυπηρετηθεί. Κάτι τέτοιο συμβαίνει στην περίπτωση του M/M/1/2, σε αντίθεση με τα M/M/1/1 και M/M/1/2*, όπου συμπεριφέρονται με τον ίδιο τρόπο στο $\lambda \rightarrow \infty$, καθώς εξυπηρετούνται τα πιο πρόσφατα πακέτα.

4.2.2 Σύγκριση M/M/1/2* και M/M/1

Η χρησιμότητα του packet management, μπορεί να φανεί και από το Σχήμα 4.8, με τη σύγκριση του M/M/1 (όπου δεν έχουμε packet management) με το M/M/1/2*.



Σχήμα 4.8 - Η μέση ηλικία της πληροφορίας σε συνάρτηση με το ρυθμό αφίξεων λ για τα μοντέλα M/M/1, M/M/1/2* για σταθερό $\mu=1$ [2]

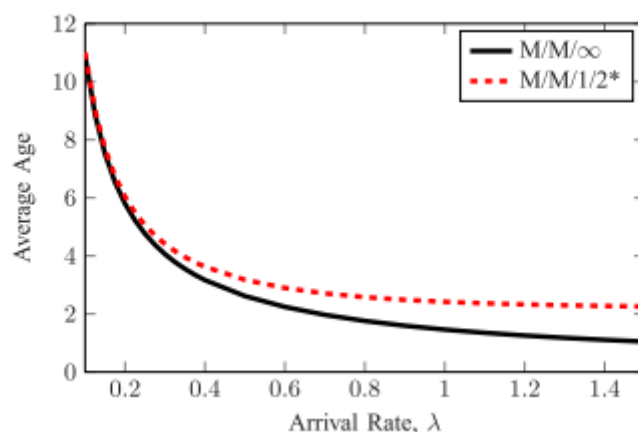
Παρατηρούμε πως για χαμηλές τιμές του λ όπου ($\lambda \ll \mu=1$), τα δύο μοντέλα έχουν παρόμοια συμπεριφορά, καθώς οι αφίξεις δεν είναι τόσο συνεχείς, με αποτέλεσμα να υπάρχει αυξημένη μέση ηλικία. Όσο αυξάνεται το λ προσεγγίζοντας την τιμή του ρυθμού εξυπηρέτησης $\mu=1$, βλέπουμε μια συνεχής μείωση της average age για το μοντέλο M/M/1/2*, όπως και προηγουμένως.

Ωστόσο, για το M/M/1 δεν ισχύει το ίδιο, αφού παρά την αρχική μείωση της μέσης ηλικίας μέχρι την τιμή $\lambda = \frac{\mu}{2}$, στη συνέχεια πραγματοποιείται απότομη αύξηση, καθώς $\lambda \rightarrow \mu$. Το τελευταίο είναι αναμενόμενο, δεδομένου ότι τα πακέτα θα μένουν στην

αναμονή για πάρα πολύ χρόνο έως ότου εξυπηρετηθούν, “κουβαλώντας” έτσι πιο παλιά πληροφορία.

4.2.3 Σύγκριση $M/M/1/2^*$ και $M/M/\infty$

Ενδιαφέρουσα είναι και η σύγκριση για τα μοντέλα $M/M/\infty$, $M/M/1/2^*$ όπου αποτυπώνεται στο Σχήμα 4.9.



Σχήμα 4.9 - Η μέση ηλικία της πληροφορίας σε συνάρτηση με το ρυθμό αφίξεων λ για τα μοντέλα $M/M/\infty$, $M/M/1/2^*$ για σταθερό $\mu=1$ [2]

Βλέπουμε πως το μοντέλο $M/M/\infty$ μας δίνει χαμηλότερες τιμές για την μέση ηλικία της πληροφορίας απ’ ότι το μοντέλο $M/M/1/2^*$. Η διαφορά, μάλιστα, στις τιμές των δύο γίνεται μεγαλύτερη καθώς ο ρυθμός αφίξεων λ αυξάνεται, αφού το $M/M/\infty$ έχει άπειρους εξυπηρετητές με αποτέλεσμα τα πακέτα να μην περιμένουν στην ουρά. Δηλαδή, η άμεση αυτή εξυπηρέτηση των πακέτων οδηγεί σε μικρότερες τιμές AoI, αφού γίνονται συνεχή status updates. Ωστόσο, δεν είναι απαραίτητα πιο ωφέλιμο, λαμβάνοντας υπόψιν την πολύ μεγάλη σπατάλη δικτυακών πόρων που συμβαίνει στο $M/M/\infty$, με τη διαθεσιμότητα τόσων εξυπηρετητών καθώς και τη μεταφορά αρκετών μη πληροφοριακών (non-informative) πακέτων.

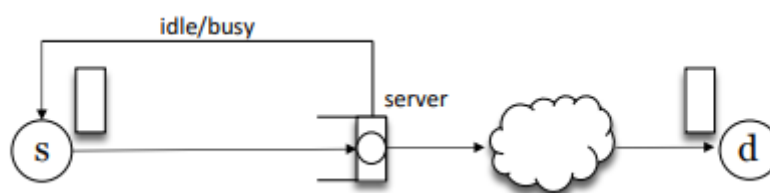
4.3 Σύγκριση Διαχείρισης Πακέτων και LCFS

Τόσο η χρήση του packet management όσο και η χρήση της πολιτικής LCFS έχει ως σκοπό τη μείωση της ηλικίας της πληροφορίας. Κατά τη LCFS όμως, αποφεύγεται η καθυστέρηση των νέων πακέτων στην αναμονή της ουράς, διότι εκείνα έχουν εντολή να πηγαίνουν πρώτα, με αποτέλεσμα σημαντική μείωση τόσο της average AoI, όσο και της PAoI (ειδικά στη LCFS with preemption). Αυτό δε συμβαίνει στις πολιτικές του packet management όπου έχουμε μεγαλύτερη καθυστέρηση στην ουρά.

Επίσης, με τη LCFS έχουμε καλύτερη αντιμετώπιση όταν υπάρχουν errors κατά τη μετάδοση των πακέτων, όπως να χαθούν λόγω της πεπερασμένης ουράς, αφού μεταδίδει αποτελεσματικά τα informative πακέτα εξασφαλίζοντας καλύτερο ρυθμό παράδοσης τους στον προορισμό [6].

4.4. Just-In-Time Πολιτική

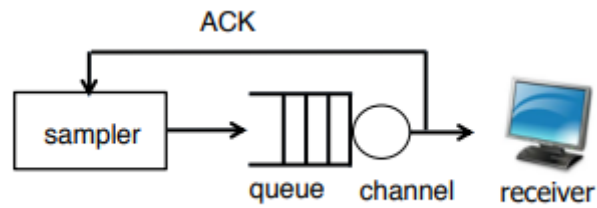
Η ελαχιστοποίηση της AoI και ο τρόπος που αυτή θα επιτευχθεί εξαρτάται από πολλούς παράγοντες όπως είναι η τοπολογία και τα χαρακτηριστικά του δικτύου, καθώς και τα διαθέσιμα μέσα για τον έλεγχό της. Ένας τρόπος για να ελέγξουμε το πόσο επίκαιρες είναι οι μεταδιδόμενες πληροφορίες είναι να υποθέσουμε πως γνωρίζουμε ακριβώς το ρυθμό εξυπηρέτησης.



Σχήμα 4.10 – Μοντέλο συστήματος με ανάδραση [9]

Θεωρούμε, δηλαδή, ότι η δημιουργία των πακέτων είναι συνάρτηση της διαδικασίας εξυπηρέτησης.

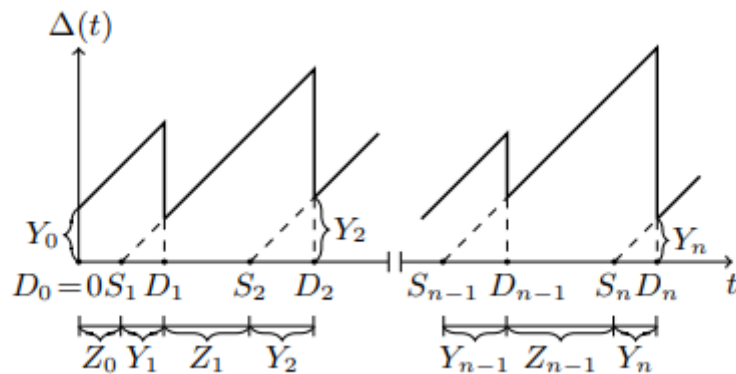
Με την πολιτική just-in-time, το update i δημιουργείται όταν τελειώσει η εξυπηρέτηση του update $i-1$, με αποτέλεσμα το πακέτο να μη χρειάζεται να περιμένει στην ουρά. Γι' αυτό ονομάζεται και zero-wait πολιτική.



Σχήμα 4.11 – Το σύστημα με ACK για ανάδραση για zero-wait πολιτική

Η πολιτική αυτή επιτυγχάνει μέγιστη ρυθμαπόδοση (throughput) και ελάχιστη καθυστέρηση (delay). Παρ' όλα αυτά δε θεωρείται η πάντα η ιδανική πολιτική σε σχέση με την AoI.

Η πηγή αντιλαμβάνεται πότε ο εξυπηρετητής είναι άδειος ή απασχολημένος μέσω των acknowledgements (ACKs). Έτσι, τα νέα updates θα δημιουργούνται όταν ο εξυπηρετητής δεν είναι απασχολημένος, δηλαδή όταν λάβει το ACK η πηγή από το κανάλι, ώστε να μην υπάρχει χρόνος αναμονής W για τα πακέτα, όπως φαίνεται στο Σχήμα 4.11.



Σχήμα 4.12 - Η μέση ηλικία της πληροφορίας σε συνάρτηση με το χρόνο για το σύστημα του σχήματος 4.10 με πολιτική Just-In-Time [15]

Θεωρούμε ότι παράγονται πακέτα τις στιγμές S_0, S_1, \dots με χρόνο μετάδοσης $Y_0, Y_1, \dots \geq 0$ του οποίου η διαδικασία Y_i είναι στατική και εργοδική Μαρκοβιανή αλυσίδα. Με $D_i = S_i + Y_i$ συμβολίζονται οι στιγμές που καταφθάνει το πακέτο i στον προορισμό ενώ με Z_i οι χρόνοι αναμονής μετά την εξυπηρέτηση του πακέτου i μέχρι να παραχθεί το επόμενο, που οφείλονται στην καθυστέρηση λήψης του ACK (Σχήμα 4.13).

Οποιαδήποτε στιγμή t , το πιο πρόσφατα ληφθέν πακέτο έχει χρονοσφραγίδα:

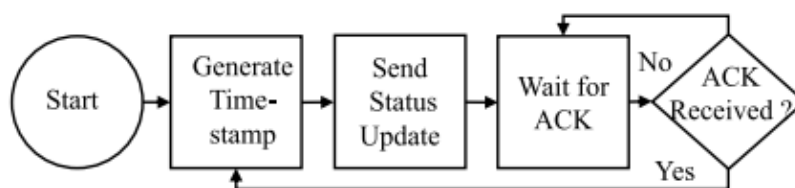
$$U(t) = \max\{S_i : D_i \leq t\} \quad (4.11)$$

Οπότε η ηλικία της πληροφορίας (Age Of Information), όπως γνωρίζουμε, υπολογίζεται από:

$$\Delta(t) = t - U(t) \quad (4.12)$$

Υποθέτουμε ότι για το πακέτο 0, παράγεται και στέλνεται τη στιγμή $S_0 = -Y_0$ έτσι ώστε να φτάνει στον προορισμό την $D_0 = 0$ (Σχήμα 4.12), με αποτέλεσμα η ηλικία της πληροφορίας να παίρνει την τιμή $\Delta(D_0) = Y_0$. Έτσι τη στιγμή D_i όπου λαμβάνεται το πακέτο i , έχουμε ηλικία $\Delta(D_i) = Y_i$ ενώ τη στιγμή S_{i+1} όπου στέλνεται το πακέτο $i+1$, έχει εισαχθεί χρόνος αναμονής $Z_i = S_{i+1} - D_i$, λόγω του ACK, και έχουμε $\Delta(S_{i+1}) = Y_i + Z_i$.

Στη συνέχεια η ηλικία συνεχίζει να αυξάνεται γραμμικά και να φτάνει την τοπικά μέγιστη τιμή της οριακά πριν ληφθεί το πακέτο $i+1$ από τον προορισμό (D_{i+1}), όπου είναι $\Delta(D_{i+1}) = Y_i + Z_i + Y_{i+1}$. Με την άφιξη του πακέτου $i+1$ στον προορισμό, η ηλικία πέφτει απευθείας αφού εκείνο έχει χρονοσφραγίδα $U_{i+1}(t) = S_{i+1}$, οπότε έχουμε $\Delta(D_{i+1}) = D_{i+1} - S_{i+1} = Y_{i+1}$.



Σχήμα 4.13 – Διάγραμμα ροής για τη Zero-Wait πολιτική [16]

Η πολιτική αυτή, όμως, δεν οδηγεί πάντα στην ελαχιστοποίηση του Age Of Information. Για παράδειγμα, το γεγονός ότι στέλνονται τα updates χωρίς να υπάρχει καθυστέρηση, μπορεί να λειτουργήσει μη ευεργετικά, σε περίπτωση συνεχόμενων πακέτων με μηδενικούς χρόνους μετάδοσης. Συγκεκριμένα, τα updates μετά από το πρώτο δε φέρνουν κάποια νέα πληροφορία στον προορισμό αφού τα συνεχόμενα

πακέτα δημιουργούνται την ίδια στιγμή, αν και η καθυστέρηση λόγω του ACK είναι μηδενική. Αυτό συμβαίνει γιατί κατά τη zero wait πολιτική, το επόμενο πακέτο δημιουργείται και στέλνεται όταν υπάρχει ενημέρωση ότι τελείωσε το προηγούμενο. Δηλαδή, αν έχουμε χρόνους μετάδοσης $Y_1, Y_2, Y_3 = 0$ για τρία συνεχόμενα updates με το πρώτο να δημιουργείται τη στιγμή $S_1 = 0$, εκείνο θα φτάσει στον προορισμό την $D_1 = 0$. Αντίστοιχα τα updates 2,3 θα δημιουργηθούν, θα σταλούν και θα φθάσουν στον προορισμό τη στιγμή 0, με αποτέλεσμα να σπαταληθούν αφού δε θα φέρουν νέα πληροφορία.

Κεφάλαιο 5: Μοντέλα ΑοΙ Διακριτού Χρόνου

5.1 Εισαγωγή στα Διακριτά Μοντέλα

Όπως έχουμε πει η ηλικία της πληροφορίας δίνεται από τον τύπο:

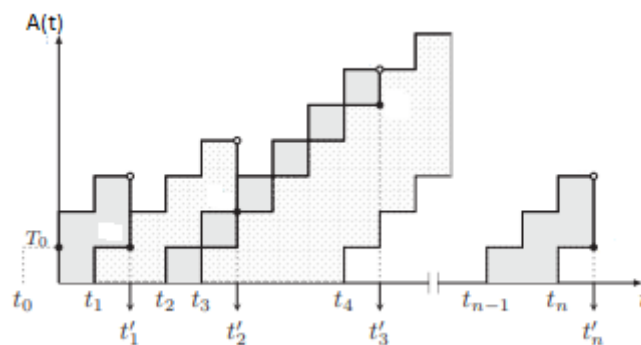
$$\Delta(t) = t - u(t)$$

Η τυχαία διαδικασία $\Delta(t)$ μπορεί να είναι και διακριτή, εκτός από συνεχής.

Θεωρούμε ένα σύστημα με πακέτα που παράγονται από μία πηγή σύμφωνα με τυχαία διαδικασία. Το σύστημα εξυπηρέτησης, που μπορεί να περιέχει έναν ή περισσότερους εξυπηρετητές, εξυπηρετεί τα πακέτα σε τυχαία ακέραια time slots. Για λόγους ανάλυσης, υποθέτουμε πως η άφιξη των πακέτων γίνεται στην αρχή του time slot ενώ η εξυπηρέτηση στο τέλος του.

Παίρνουμε τη διαδικασία $A(t)$ που περιγράφει την ηλικία, από την τιμή που έχει η ΑοΙ στην αρχή κάθε time slot. Υποθέτοντας ότι το i -οστό πακέτο δημιουργείται τη στιγμή Y_i , ισχύει:

$$A(t+1) = \begin{cases} A(t) + 1, & \text{αν δεν έχει γίνει εξυπηρέτηση τη στιγμή } t \\ \min\{t - Y_i, A(t)\} + 1, & \text{αν το πακέτο } i \text{ έχει εξυπηρευτεί} \end{cases} \quad (5.1)$$



Σχήμα 5.1 - Η ηλικία της πληροφορίας της διακριτής $A(t)$

Για την average και peak age αντίστοιχα ισχύει από [10]:

$$A_{ave} = \lim_{T \rightarrow \infty} \sup \frac{1}{T} \sum_{t=1}^T A(t) \quad (5.2)$$

και

$$A_p = \lim_{T \rightarrow \infty} \sup \frac{\sum_{t=1}^{t=T} A(t) \mathbb{1}\{A(t+1) \leq A(t)\}}{\sum_{t=1}^{t=T} \mathbb{1}\{A(t+1) \leq A(t)\}} \quad (5.3)$$

Παρακάτω παρουσιάζονται διάφορα διακριτά μοντέλα μαζί με τα απαραίτητα metrics για την ηλικία της πληροφορίας.

5.2 First Come First Served Ουρές

Θεωρούμε μία $G/G/1$ ουρά. Τα πακέτα δημιουργούνται ανάμεσα σε χρόνους X (inter-generation times) που ακολουθούν κατανομή F_x ενώ οι χρόνοι εξυπηρέτησης S είναι ομοιόμορφες ανεξάρτητες μεταβλητές που ακολουθούν την F_s .

Έτσι, προκύπτουν: $\lambda = \frac{1}{E[X]}$ και $\mu = \frac{1}{E[S]}$.

Αποδεικνύεται (στο [11]) πως το κάτω όριο της μέσης ηλικίας για το $G/G/1$ μοντέλο είναι:

$$A_{ave L, G/G/1} = \frac{1}{2\lambda} + \frac{1}{\mu} \quad (5.4)$$

αφού $A_{ave L, G/G/1} \geq \frac{1}{2\lambda} + \frac{1}{\mu}$

Αυτό επιτυγχάνεται όταν η δημιουργία των πακέτων είναι περιοδική και ο χρόνος εξυπηρέτησης ντετερμινιστικός. Γενικότερα, ο ντετερμινισμός στη δημιουργία και την εξυπηρέτηση ελαχιστοποιεί τα δύο metrics που είδαμε παραπάνω, δηλαδή την average και την peak age.

Δηλαδή, σε ένα FCFS σύστημα ισχύει:

$$\left\{ \begin{array}{l} A^{\text{ave}}_{D/D/1} \leq A^{\text{ave}}_{G/G/1} \\ A^p_{D/D/1} \leq A^p_{G/G/1} \end{array} \right. \quad (5.5)$$

για οποιοδήποτε ρυθμό δημιουργίας και εξυπηρέτησης πακέτων.

Το παραπάνω συμπέρασμα μας οδηγεί να εξετάσουμε και ορισμένα μοντέλα ώστε να δούμε τα metrics τους που εξαρτώνται από τις κατανομές των X, S .

5.2.1 $G/M/1$ – *First Come First Served*

Θεωρούμε μια ουρά $G/M/1$, όπου η κατανομή F_S του χρόνου εξυπηρέτησης είναι εκθετική με ρυθμό μ και ισχύει: $F_S(x) = 1 - e^{-\mu x}$. Οι χρόνοι δημιουργίας X των πακέτων είναι γενικά κατανεμημένοι σύμφωνα με την F_x , με μέσο $\frac{1}{\lambda}$.

Υποθέτοντας $M_x = E[e^{\alpha x}]$

και

$$\bar{\alpha} \text{ ως τη μοναδική λύση της } \alpha = \mu - \mu M_x(-\alpha)$$

Η μέση ηλικία της πληροφορίας είναι από [11]:

$$A_{\text{ave}, G/M/1} = \lambda \left[\frac{1}{2} M''_x(0) + \frac{1}{\bar{\alpha}} M'_x(-\bar{\alpha}) \right] + \frac{1}{\mu} \quad (5.6)$$

ενώ η μέγιστη:

$$A_{p, G/M/1} = \frac{1}{\bar{\alpha}} + \frac{1}{\lambda} \quad (5.7)$$

5.2.2 $M/G/1$ – *First Come First Served*

Θεωρούμε μια ουρά $G/M/1$, όπου η κατανομή F_x του χρόνου δημιουργίας των πακέτων είναι Poisson με ρυθμό λ και ισχύει: $F_x(x) = 1 - e^{-\lambda x}$. Οι χρόνοι εξυπηρέτησης S των πακέτων είναι γενικά κατανεμημένοι σύμφωνα με την F_S , με μέσο $\frac{1}{\mu}$.

Οι σχέσεις για τις μετρικές σε αυτό το μοντέλο είναι [11]:

$$A_{ave,M/G/1} = \frac{1}{\mu} \left[1 + \left(\frac{\rho}{1-\rho} \right) \frac{E[S^2]}{2E[S]^2} + \left(\frac{1-\rho}{\rho} \right) \frac{1}{E[e^{-\lambda S}]} \right] \quad (5.8)$$

και

$$A_{p,M/G/1} = \frac{1}{\mu} \left[1 + \frac{1}{\rho} + \frac{\rho}{1-\rho} \frac{E[S^2]}{2E[S]^2} \right] \quad (5.9)$$

5.2.3 BER/G/1 – First Come First Served

Θεωρούμε μια διακριτού χρόνου ουρά Ber/G/1, με αφίξεις Bernoulli που γίνονται στιγμές t με πιθανότητα λ . Οι χρόνοι εξυπηρέτησης είναι γενικά κατανομημένοι με μέσο $E[S] = \frac{1}{\mu}$.

Η μέση ηλικία για την BER/G/1 [10] είναι :

$$A_{ave,BER/G/1} = 1 + \frac{1}{\mu} + \frac{(1-\lambda)(1-\rho)}{\lambda L_S (1-\lambda)} + \frac{\lambda E[S^2] - \rho}{2(1-\rho)} \quad (5.10)$$

ενώ η μέγιστη ηλικία:

$$A_{p,BER/G/1} = \frac{1}{\lambda} + \frac{1}{\mu} + \frac{\lambda E[S^2] - \rho}{2(1-\rho)} \quad (5.11)$$

όπου $L_S(x) = E[x^S]$ είναι η συνάρτηση που περιγράφει τη δημιουργία πιθανότητας.

5.2.4 BER/G/1 with vacations – First Come First Served

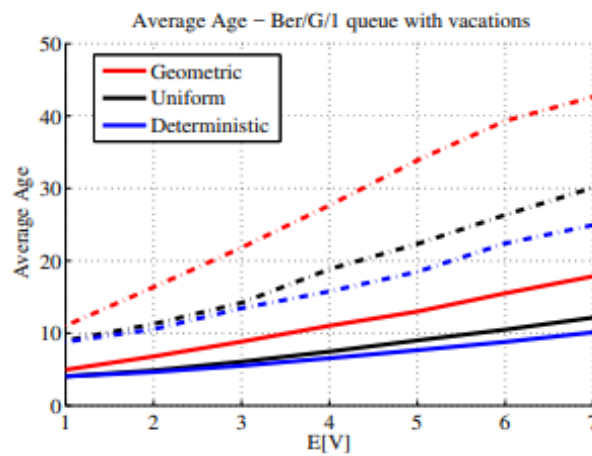
Θεωρούμε όμοια με πριν μια Ber/G/1 ουρά με τη διαφορά πως όταν η ουρά είναι κενή, ο εξυπηρετητής κάνει διακοπές σε χρόνους που περιγράφονται από τη V , ομοιόμορφα και ανεξάρτητα κατανομημένους με μέσο $E[V]$, έως ότου γίνει μια νέα άφιξη στην ουρά.

Η μέγιστη ηλικία όπως περιγράφεται στο [10], είναι:

$$A_{p,BER/G/1,VAC} = \frac{1}{\lambda} + \frac{1}{\mu} + \frac{\lambda E[S^2] - \rho}{2(1-\rho)} + \frac{E[V^2]}{2E[V]} - \frac{1}{2} \quad (5.12)$$

ενώ η μέση ηλικία προσδιορίζεται από το ανώτατο όριο (upper bound) που είναι η μέγιστη ηλικία, δηλαδή:

$$A_{ave,BER/G/1,VAC} \leq A_{p,BER/G/1,VAC} \quad (5.13)$$



Σχήμα 5.2 - Η μέση ηλικία της πληροφορίας σε συνάρτηση με το μέσο χρόνο διακοπών για το μοντέλο BER/G/1 με διακοπές για σταθερό $\mu=0.75$ και εξυπηρετήσεις γεωμετρικά κατανομημένες (BER/Geo/1). Η κατανομή του V μπορεί να είναι Γεωμετρική, Ομοιόμορφη και Ντετερμινιστική. Οι συνεχείς γραμμές αναπαριστούν πιθανότητα αφίξεως $\lambda=0.3$ ενώ οι διακεκομμένες $\lambda=0.6$. [10]

Από το Σχήμα 5.2 για μια δεδομένη τιμή του $E[V]$ βλέπουμε πως η ντετερμινιστική κατανομή του V μας δίνει χαμηλότερες τιμές για τη μέση ηλικία σε σύγκριση με τη γεωμετρική και την ομοιόμορφη. Επίσης, επιτυγχάνεται μικρότερη μέση ηλικία για χαμηλότερη πιθανότητα αφίξεως $\lambda=0.3$ απ' ό τι για $\lambda=0.6$

5.2.5 G/BER/1 – First Come First Served

Θεωρούμε μια διακριτού χρόνου ουρά G/BER/1, με χρόνους δημιουργίας X των πακέτων είναι γενικά κατανομημένοι σύμφωνα με την F_x , με μέσο $\frac{1}{\lambda}$. Η εξυπηρέτηση γίνεται σύμφωνα με την κατανομή Bernoulli με ρυθμό μ .

Υποθέτοντας $M_x = E[e^{\alpha x}]$

και

α^* ως τη μοναδική λύση της $\alpha = \mu - \mu M_x(\log(1-\alpha))$

Από [12] προκύπτει ότι:

Η μέση ηλικία της πληροφορίας είναι :

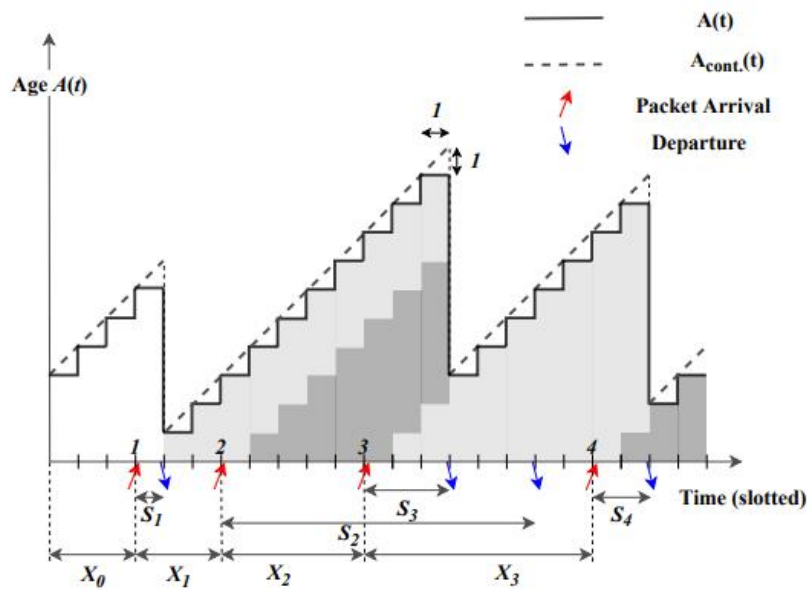
$$A_{ave,G/BER/1} = \lambda \left[\frac{1}{2} M''x(0) + \frac{1}{a^*} M'x(\log(1 - a^*)) \right] + \frac{1}{\mu} + \frac{1}{2} \quad (5.14)$$

ενώ η μέγιστη:

$$A_{p,G/BER/1} = \frac{1}{a^*} + \frac{1}{\lambda} \quad (5.15)$$

5.3 Συσχέτιση Διακριτής με Συνεχή AoI

Σε προηγούμενα κεφάλαια έχουμε δει πως προκύπτουν οι τιμές για την ηλικία της πληροφορίας σε συνεχούς χρόνου ουρές, ενώ παραπάνω εξετάσαμε τις αντίστοιχες διακριτές. Το ερώτημα που υπάρχει είναι πως εκείνες συσχετίζονται.



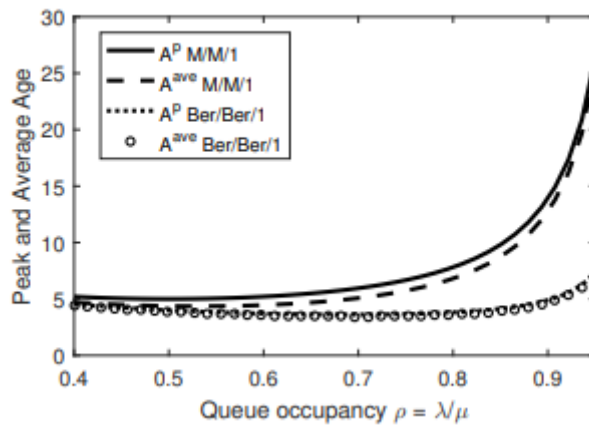
Σχήμα 5.3 - Η ηλικία της πληροφορίας της διακριτής $A(t)$ σε συνάρτηση με το χρόνο μαζί με την αντίστοιχη συνεχή διαδικασία $A_{cont}(t)$ [13]

Δεδομένου ότι οι σχέσεις που μας δίνουν τη μέγιστη και τη μέση ηλικία της πληροφορίας, για τη συνεχούς χρόνου διαδικασία, έχει οριστεί με ακρίβεια, οι αντίστοιχες διακριτές δίνονται από τις σχέσεις:

$$\begin{cases} A^{p_{discr}} = A^{p_{cont}} - 1 \\ A^{ave_{discr}} = A^{ave_{cont}} - 1 \end{cases} \quad (5.16)$$

5.3.1 Σύγκριση Συνεχούς M/M/1 με Διακριτή BER/BER/1

Παρακάτω, παρουσιάζεται η peak και average ηλικία σε συνάρτηση με τη χρησιμοποίηση ρ για τη συνεχές M/M/1 και τη διακριτή BER/BER/1.



Σχήμα 5.4 - Η μέγιστη και η μέση ηλικία της πληροφορίας σε σχέση με τη χρησιμοποίηση ρ για τη συνεχές M/M/1 και τη διακριτή BER/BER/1 για $\mu=0.8$. [12]

Από το Σχήμα 5.4, βλέπουμε πως οι τιμές για τη διακριτή BER/BER/1 είναι αρκετά χαμηλότερες τόσο για τη μέγιστη όσο και για τη μέση ηλικία τη πληροφορίας. Ειδικότερα, για καθώς το ρ πλησιάζει το 1, η τιμές των $A^P_{M/M/1}$ και $A^{ave}_{M/M/1}$ αυξάνονται αρκετά, κάνοντας μεγαλύτερη τη διαφορά μεταξύ εκείνων του BER/BER/1.

Κεφάλαιο 6: Age of Information στα Κοινωνικά Δίκτυα και APIs

6.1 Κοινωνικά Δίκτυα

Τα κοινωνικά δίκτυα είναι ιστοσελίδες που επιτρέπουν τη διεπαφή μεταξύ των χρηστών τους μέσω διαφόρων αλληλεπιδράσεων. Αποτελούν εικονικές κοινότητες μέσα από τις οποίες οι χρήστες μπορούν να μοιραστούν ιδέες, φωτογραφίες, βίντεο καθώς και να μιλήσουν μεταξύ τους εκφραζόμενοι με πολλούς τρόπους.

Υπάρχουν ποικίλα κοινωνικά δίκτυα, που συνεχώς αναπτύσσουν και βελτιώνουν τη λειτουργία τους προσφέροντας ολοένα και καλύτερο περιβάλλον στους χρήστες τους και διευκολύνοντας την αλληλεπίδραση μεταξύ τους. Μερικά από τα πιο γνωστά είναι τα Facebook, Twitter, Instagram, LinkedIn.

6.2 Κοινωνική Ηλικία της Πληροφορίας (Social Age of Information)

Μέχρι στιγμής έχουμε δει πως ορίζεται η Ηλικία της Πληροφορίας (Age of Information) την οποία και εξετάσαμε για τα διάφορα συνεχή και διακριτά μοντέλα. Για τα κοινωνικά δίκτυα, μπορούμε να ορίσουμε μια μετρική, που θα παρατηρεί πόσο πρόσφατη είναι μια πληροφορία. Έτσι, εισάγουμε τη Social Age of Information, που μετράει πόσος χρόνος έχει περάσει από την τελευταία δραστηριότητα που έχει επιτευχθεί σε ένα κοινωνικό δίκτυο. Η πληροφορία αυτή μπορεί να ορίζεται ως like, comment, post και οποιαδήποτε άλλη δραστηριότητα μπορεί να γίνει σε ένα συγκεκριμένο πλαίσιο που επιδέχεται κοινωνικές διαδικτυακές ενέργειες. Για παράδειγμα, μπορεί να μετρηθεί πόσο πρόσφατη είναι η τελευταία δημοσίευση σε ένα timeline ενός χρήστη, ή το τελευταίο comment σε ένα post. Η συγκεκριμένη μετρική αποτελεί συμβολή της παρούσας εργασίας.

Αν θεωρήσουμε ως u τη στιγμή που έλαβε πράξη η τελευταία δραστηριότητα (δηλαδή η αντίστοιχη χρονοσφραγίδα της), τότε μια τυχαία στιγμή $t > u$, η Social Age of Information είναι:

$$S(t) = t - u \quad (6.1)$$

Μεταξύ των διαδοχικών πληροφοριών που δίνουν οι χρήστες, μεσολαβούν κάποια χρονικά διαστήματα, τα οποία παίρνουν μέγιστη τιμή τη στιγμή πριν δημοσιευτεί η επόμενη δραστηριότητα. Αυτή ορίζεται και ως peak Social Age of Information και αναπαριστά το χρόνο που μεσολαβεί μεταξύ των διαδοχικών «αντιδράσεων».

Πως είναι δυνατόν όμως να αντληθούν αυτά τα δεδομένα από ένα κοινωνικό δίκτυο; Αυτό οδηγεί στην αναζήτηση της χρησιμότητας των API's.

6.3 APIs

6.3.1 Εισαγωγή στα APIs

Το API (Application Programming Interface) είναι Διεπαφή Προγραμματισμού Εφαρμογών ή αλλιώς Διασύνδεση Προγραμματισμού Εφαρμογών. Πιο συγκεκριμένα, είναι ένα ενδιάμεσο λογισμικό που παρέχει ένα λειτουργικό σύστημα, βιβλιοθήκη ή εφαρμογή προκειμένου να επιτρέπει να γίνονται προς αυτά αιτήσεις από άλλα προγράμματα ή/και ανταλλαγή δεδομένων [17].

Ένα API αποτελείται συνήθως από ένα σύνολο ρουτινών, πρωτοκόλλων και εργαλείων που χρησιμοποιούνται για την κατασκευή εφαρμογών λογισμικού. Εν ολίγοις, καθορίζει πώς πρέπει να ενσωματωθεί το λογισμικό και πώς λειτουργεί αυτό το λογισμικό στη συνέχεια. Παραδίδει το αίτημά του χρήστη στον πάροχο που ανήκει ο τελευταίος, και στη συνέχεια επιστέφει την απάντηση πίσω στον χρήστη. Να σημειωθεί πως, για όσο ο πάροχος υπηρεσιών παραδίδει απαντήσεις στο τελικό σημείο, οι μεταβολές στην υποδομή πίσω από το τελικό σημείο δεν είναι ορατές από τις εφαρμογές που βασίζονται στο συγκεκριμένο API. Ως εκ τούτου, ο πάροχος υπηρεσιών δίνει μεγάλη ευελιξία σε ό,τι αφορά τις προσφερόμενες υπηρεσίες.

Τα περισσότερα λειτουργικά περιβάλλοντα παρέχουν API όπως και οι περισσότεροι μεγάλοι ιστότοποι αυτές τις μέρες. Αυτά τα API παρέχονται για να διευκολύνουν τους προγραμματιστές να συνδέσουν το δικό τους λογισμικό με το λειτουργικό σύστημα ή τον ιστότοπο.

6.3.2 Social Networks APIs

Για να αντλήσουμε μαζικά τα δεδομένα μιας δραστηριότητας από τα κοινωνικά δίκτυα ώστε να μπορούμε να τα επεξεργαστούμε και να τα αναλύσουμε χρησιμοποιούμε τα API's τους.

Παρ' ότι τα κοινωνικά δίκτυα ήταν αρχικά σχεδιασμένα ώστε να συγκρατούν την εμπειρία του χρήστη σε αυτά που προσφέρει η ιστοσελίδα τους, δημιουργήθηκε η ανάγκη να επεκτείνουν τη λειτουργία τους παρέχοντας δικά τους API's. Τα τελευταία επιτρέπουν σε άλλες εφαρμογές και ιστότοπους να αντλούν τα δεδομένα των μέσω κοινωνικής δικτύωσης και να ενσωματώνονται στον ιστότοπο ή την εφαρμογή τους. Έτσι, δίνουν τη δυνατότητα σε third-party developers να συνδέσουν το δικό τους λογισμικό με το κοινωνικό δίκτυο και να αναπτύξουν εφαρμογές που μπορούν να βελτιώσουν τη χρήση και τη λειτουργία της κεντρικής πλατφόρμας.

Βέβαια, παρέχουν περιορισμένες δυνατότητες, κλήσεις ή ερωτήματα που μπορούν να εκτελεστούν από τον προγραμματιστή και απαιτούν να δημιουργήσει ένα κλειδί API πριν από τη χρήση του.

6.3.3 Facebook API

Γενικά, το Facebook είναι λιγότερο ανοιχτό σε σχέση με μερικά άλλα κοινωνικά δίκτυα, όπως το Twitter, όσον αφορά την κοινή χρήση πληροφοριών. Το απόρρητο είναι πιο αυστηρό επειδή, κατά βάση, οι άνθρωποι μοιράζονται περισσότερα προσωπικά δεδομένα στο Facebook.

Είναι δυνατόν να χρησιμοποιήσουμε τα API του Facebook για να εξάγουμε δεδομένα σχετικά με άτομα. Το τι και πόσα δεδομένα μπορεί να εξάγει κανείς, εξαρτάται από τις ρυθμίσεις απορρήτου τους, τα προνόμια που έχουν παραχωρήσει στην εκάστοτε εφαρμογή κλπ. Αν και είναι κάπως αμφισβητήσιμο και μάλλον αντίκειται στους όρους χρήσης του Facebook, είναι τεχνικά δυνατό να δημιουργηθεί ένας λογαριασμός και να χρησιμοποιηθεί ένα Token OAuth για τη συλλογή δεδομένων σχετικά με άτομα που το έχουν (αν και ο όγκος των δεδομένων που μπορεί να συλλεχθεί, πάλι, εξαρτάται από τις ρυθμίσεις απορρήτου τους).

Τα Token OAuth είναι ειδικά για τον χρήστη. Όταν ένας χρήστης συνδέεται και εξουσιοδοτεί μια εφαρμογή να έχει πρόσβαση στον λογαριασμό του, δίνεται ένα token που μπορεί να χρησιμοποιηθεί για αυτόν τον σκοπό. Μπορούν να πραγματοποιηθούν ανώνυμα αιτήματα Graph API, καθώς δεν απαιτείται σε κάθε αίτημα access token. Ωστόσο, τα δεδομένα σχετικά με τους χρήστες περιορίζονται σε αυτή την περίπτωση.

Τα τελευταία χρόνια έχουν αυστηροποιηθεί ακόμα περισσότερο οι άδειες που δίνονται για «άντληση» πληροφοριών από το Facebook API. Ένας χρήστης μπορεί να κάνει αίτημα μόνο για να πάρει τα δικά του likes ή comments, λόγω ιδιωτικότητας και προστασίας των προσωπικών δεδομένων. Ειδικότερα χρησιμοποιώντας το user access token, επιστρέφονται κενά δεδομένα για διαφόρους κόμβους.

Παρακάτω παρουσιάζεται ο κώδικας με τον οποίο μπορούμε να εκμαιεύσουμε από το Facebook τις προσωπικές μας πληροφορίες (σε αρχείο txt) όπως τα likes που έχουμε κάνει σε σελίδες καθώς και τη ροή ειδήσεων με τα αντίστοιχα timestamps. Παρ' όλα αυτά, όπως είπαμε αυτά αφορούν τα ατομικά δεδομένα.

```
import facebook
import datetime

token = 'an access token'
user_id = 'the id of the user'

def get_likes_api ():
    try:
        graph = facebook.GraphAPI(access_token = token , version=3.1)
        likes = graph.request ('/user_id/likes') ['data']
        user = graph.request ('/user_id?fields=name')

        newlist = []
        timelist = []
        for dic in likes:
            newlist.append(dic['name'])
            timelist.append(dic['created_time'])

        finallist = list(zip(newlist,timelist))

        if newlist !=[]:
            now = datetime.datetime.now()
            filename = str(user['name']) + '-likes-on' +
            now.strftime ('-%y-%m-%d-' + '@-%H-%M')

            with open(filename + '.txt', 'w') as fp:
                fp.write('\n'.join('%s %s' % x for x in finallist))
```

```

        print('likes successfully retrieved.')

    else:
        print('There were no likes.')

except Exception as e:
    print(e)

def get_feed_api ():
    try:
        graph = facebook.GraphAPI(access_token = token , version=3.1)
        feed = graph.request ('/user_id/feed') ['data']
        user = graph.request ('/user_id?fields=name')

        newlist = []
        timelist = []
        for dic in feed:
            newlist.append(dic['id'])
            timelist.append(dic['created_time'])

        finallist = list(zip(newlist,timelist))

        if newlist !=[]:
            now = datetime.datetime.now()
            filename = str(user['name']) + '-feed-on' +
            now.strftime('-%y-%m-%d-' + '@-%H-%M')

            with open(filename + '.txt', 'w') as fp:
                fp.write('\n'.join('%s %s' % x for x in finallist))

            print('Feed successfully retrieved.')

        else:
            print('There was no posts in feed.')

    except Exception as e:
        print(e)

```

Κώδικας 6.1 – Εξαγωγή ατομικών πληροφοριών από Facebook API.

Κεφάλαιο 7: Παρουσίαση Dataset και Data Analysis

7.1 Άντληση και Επεξεργασία Δεδομένων από το Dataset

Δεδομένου της αδυναμίας, λόγω του αυστηρού απορρήτου, να αντλήσουμε την ποσότητα αλλά και την ποιότητα δεδομένων που χρειαζόμαστε μέσω του Facebook API, θα χρησιμοποιηθεί ένα έτοιμο dataset που περιέχει μεταξύ άλλων την ακριβή ημερομηνία και ώρα των comments διαφόρων post.

Το dataset αυτό προέρχεται από [18].

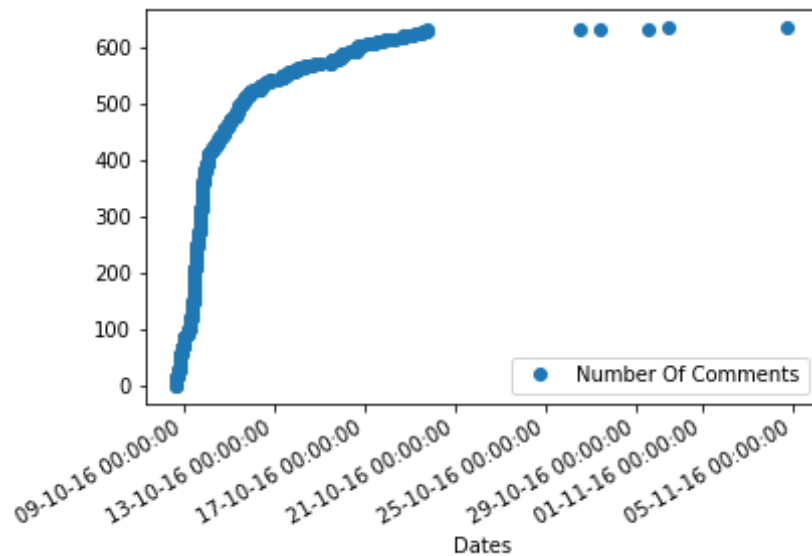
Εμείς επιλέγουμε ένα post με status id : 644891892279936_803410486428075 το οποίο και επεξεργαζόμαστε για να χρησιμοποιηθεί σωστά για τις ανάγκες της μελέτης που θα γίνει. Συγκεκριμένα, περιέχει 636 comments τα οποία και ταξινομούμε με αυξανόμενη σειρά σχετικά με το πότε δημιουργήθηκαν.

Επίσης, υπολογίζουμε τους χρόνους (σε seconds) μεταξύ των διαδοχικών comments καθώς και το χρόνο που έχει περάσει από το πρώτο comment (σε seconds) όπου θεωρείται η στιγμή μηδέν.

#	A	B	C	D	E	F	G	H	
1	comment_id	status_id	parent_id	comment_mes	comment_author	comment_published	social_peak	aoi	time
2	803410486428075_803414059	644891892279936_803410486428075		THIS IS EXACTI	Debbie Taylor	2016-10-08 15:13:10	0		0
3	803410486428075_803419226	644891892279936_803410486428075		- I'm more ofe	Paul Laskowski	2016-10-08 15:17:00	230		230
4	803410486428075_803421046	644891892279936_803410486428075		I am a strong	Shannon Smith	2016-10-08 15:21:22	262		492
5	803410486428075_803421076	644891892279936_803410486428075	803410486428075_805	And if you w	Phyllis Gore	2016-10-08 15:21:30	8		500
6	803410486428075_803424696	644891892279936_803410486428075	803410486428075_805	lol! -100% corr	Debbie Taylor	2016-10-08 15:24:01	151		651
7	803410486428075_803428699	644891892279936_803410486428075	803410486428075_805	Exactly it shov	Neal Smarra	2016-10-08 15:27:07	186		837
8	803410486428075_803432119	644891892279936_803410486428075	803410486428075_805	Monico Soto-(Shannon Smith	2016-10-08 15:29:23	136		973
9	803410486428075_803434206	644891892279936_803410486428075	803410486428075_805	Monico Soto-(Shannon Smith	2016-10-08 15:31:02	99		1072
10	803410486428075_803441489	644891892279936_803410486428075	803410486428075_805	The Clinton's	Christina Roberts Sutherland	2016-10-08 15:36:49	347		1419
11	803410486428075_803441739	644891892279936_803410486428075	803410486428075_805	Monico Soto-(Chandra Galloway	2016-10-08 15:37:30	41		1460
12	803410486428075_803442819	644891892279936_803410486428075	803410486428075_805	Monico Soto-(Olivia Bland	2016-10-08 15:41:05	215		1675
13	803410486428075_803444023	644891892279936_803410486428075	803410486428075_805	Monico Soto-(David Tetrick	2016-10-08 15:45:09	244		1919
14	803410486428075_803445766	644891892279936_803410486428075	803410486428075_805	@ Monico Sc	Paul Laskowski	2016-10-08 15:49:07	238		2157
15	803410486428075_803448553	644891892279936_803410486428075	803410486428075_805	Some do the	Maryann Rivera	2016-10-08 15:56:49	462		2619
16	803410486428075_803448993	644891892279936_803410486428075	803410486428075_805	Monico Soto-(Susan Fender	2016-10-08 15:58:01	72		2691
17	803410486428075_803451079	644891892279936_803410486428075	803410486428075_805	Look, I know	Paul Laskowski	2016-10-08 16:04:09	368		3059
18	803410486428075_803451433	644891892279936_803410486428075	803410486428075_805	Susan Fender	Nancy Crayton	2016-10-08 16:04:42	33		3092
19	803410486428075_803452106	644891892279936_803410486428075	803410486428075_805	Hey Debbie T	Kerry Oberlander	2016-10-08 16:06:45	123		3215
20	803410486428075_803454619	644891892279936_803410486428075	803410486428075_805	Right-I am w	Debbie Taylor	2016-10-08 16:10:03	198		3413
21	803410486428075_803455373	644891892279936_803410486428075	803410486428075_805	Monico Soto-(Ingrid Johnson	2016-10-08 16:11:37	94		3507
22	803410486428075_803456229	644891892279936_803410486428075	803410486428075_805	Monico Soto-(Debbie Taylor	2016-10-08 16:14:15	158		3665
23	803410486428075_803488806	644891892279936_803410486428075	803410486428075_805	Awesome! I c	Sharon Block	2016-10-08 16:46:18	1923		5588
24	803410486428075_803493719	644891892279936_803410486428075		FUK'N A!!	Anthony Jacques	2016-10-08 16:49:07	169		5757
25	803410486428075_803504633	644891892279936_803410486428075	803410486428075_805	Surgeons are	Carl Samuelson	2016-10-08 17:07:00	1073		6830
26	803410486428075_803518523	644891892279936_803410486428075	803410486428075_805	π±	Joan Kratowicz Tenaglia	2016-10-08 17:34:49	1669		8499
27	803410486428075_803520166	644891892279936_803410486428075	803410486428075_805	Hillavlic trak	Mark Stanul	2016-10-08 17:39:35	386		8785

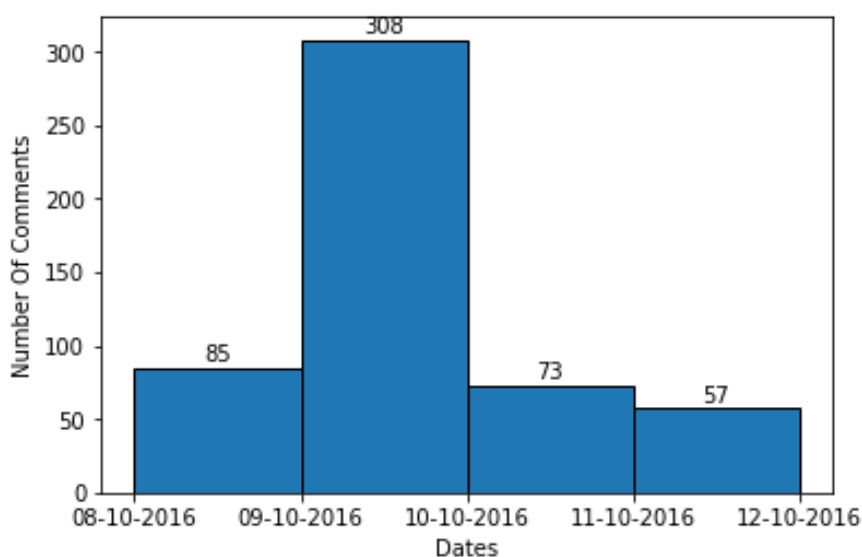
Σχήμα 7.1 – Μέρος του dataset που επεξεργάστηκε και χρησιμοποιήθηκε.

Παρακάτω βλέπουμε τον αριθμό των comments όπως προκύπτει σε σχέση με τις ημερομηνίες.



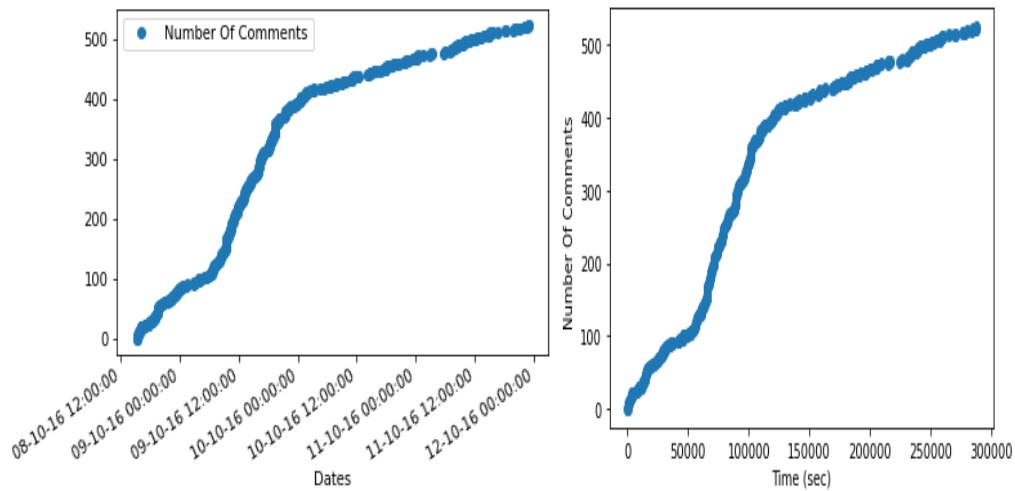
Σχήμα 7.2 – Η πορεία του αριθμού των comments στο συνολικό post σε σχέση με τις ημερομηνίες

Βλέπουμε πως η συντριπτική πλειοψηφία των comments έχει γίνει τις πρώτες μέρες, ενώ όσο περνάνε οι μέρες γίνονται ελάχιστα. Έτσι, θα ερευνήσουμε τι συμβαίνει στο post τις πρώτες 4 ημέρες (08/10 – 11/10), όπου έχουμε το μεγαλύτερο δείγμα και θα γίνει πιο ουσιαστική μελέτη με πιο ρεαλιστικά αποτελέσματα. Αυτό από εδώ και στο εξής θα θεωρούμε πως είναι το post που ερευνούμε.



Σχήμα 7.3 – Πλήθος comments στο post ανά ημέρα.

Στο διάστημα αυτό, λοιπόν, γίνονται 523 comments, όπως φαίνεται και παρακάτω:



(α)

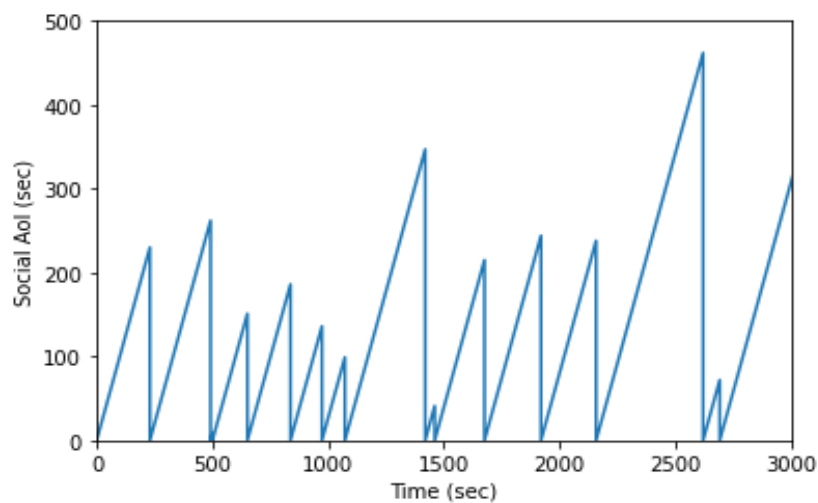
(β)

Σχήμα 7.4 – Η πορεία του αριθμού των comments για τις 4 πρώτες ημέρες (α) ως προς τις ημερομηνίες και (β) ως προς το χρόνο

7.2 Εύρεση Social Age of Information

Με βάση τους ορισμούς που δώθηκαν προηγουμένως, η Social Age of Information των comments, στο post, είναι ο χρόνος που έχει περάσει από το τελευταίο comment.

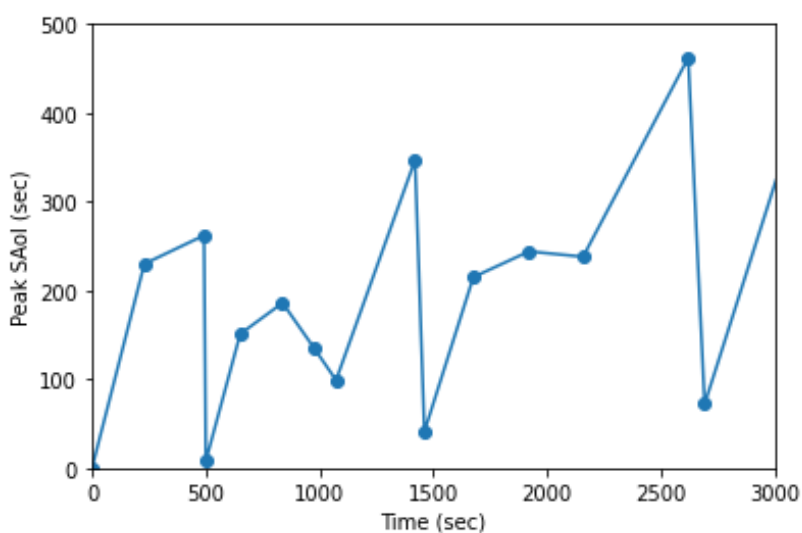
Η πορεία της για ένα μέρος του post είναι:



Σχήμα 7.5 – Η εξέλιξη της Social Age of Information σε σχέση με το χρόνο.

Η τιμή της Social AoI αρχίζει να αυξάνεται γραμμικά μέχρι τη δημοσίευση του επόμενου comment, στιγμή στην οποία και μηδενίζεται. Αυτό γίνεται για κάθε διάστημα μεταξύ των comments, παίρνοντας μέγιστη τιμή τη στιγμή ακριβώς πριν την «άφιξη» του επομένου. Στο Σχήμα 7.5, το κάθε μέγιστο παρουσιάζεται ως μια κορυφή της Social AoI πριν την απότομη κάθοδο της στο μηδέν.

Η τιμή, αυτή, προσδιορίζεται από τη μετρική που ονομάσαμε peak Social Age of Information για κάθε comment ξεχωριστά. Η μεταβλητή που την περιγράφει ορίζεται ως P και η εξέλιξη της φαίνεται στο Σχήμα 7.6:



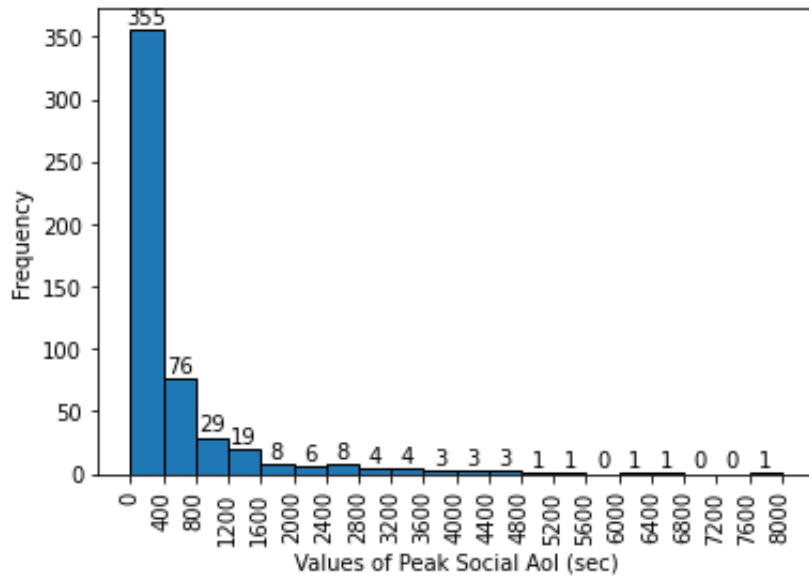
Σχήμα 7.6 – Η εξέλιξη της Peak Social Age of Information σε σχέση με το χρόνο.

Οι τιμές της peak Social AoI, είναι διακριτές και η συνεχής γραμμή εμφανίζεται για να υποδείξει την ακριβή πορεία που ακολουθούν.

7.3 Fit της Peak Social Age of Information

Θα αναζητήσουμε αν το σύνολο των τιμών της, ακολουθεί έστω και κατά προσέγγιση κάποια από τα τις γνωστές κατανομές. Επειδή, όμως, τα εργαλεία που έχουμε στην pythοn διευκολύνουν το data fitting με συνεχείς κατανομές και όχι με διακριτές, θα βρούμε τη συνεχή κατανομή που ταιριάζει καλύτερα με τις τιμές του peak Social AoI.

Πρώτα απ' όλα, παρουσιάζεται το ιστόγραμμα της :



Σχήμα 7.7 – Το ιστόγραμμα των τιμών της Peak Social Age of Information.

Είναι εμφανές από το Σχήμα 7.7 πως το μεγαλύτερο πλήθος των τιμών της P είναι κάτω από 1000 sec. Από τον τρόπο με τον οποίο κατανέμονται, θα γίνει αναζήτηση της συνεχούς κατανομής με την οποία γίνεται καλύτερο fit.

7.3.1 Fitter

Πρώτα χρησιμοποιούμε το πακέτο Fitter, με το οποίο σκανάρονται οι κατανομές που θέλουμε κάνοντας fit με αυτές. Στη συνέχεια, δίνεται μία σύνοψη σχετικά με το ποιες είναι οι καλύτερες κατανομές οι οποίες προκύπτουν από την τιμή αθροίσματος του τετραγώνου των σφαλμάτων που επιτυγχάνουν (Sum of the Square Errors - SSE). Η SSE μετράει το ποσό του σφάλματος μεταξύ της συνάρτησης της κατανομής και των δεδομένων που έχουμε, καθορίζοντας ουσιαστικά πόσο καλά γίνεται το data fitting. Προφανώς, όσο χαμηλότερη είναι αυτή η τιμή τόσο καλύτερα ταιριάζει η κατανομή στα δεδομένα.

Παρακάτω φαίνεται ο κώδικας που χρησιμοποιήθηκε σε γλώσσα προγραμματισμού python για να υλοποιηθεί το data fitting για την peak Social AoI:


```

from fitter import Fitter
import pandas as pd

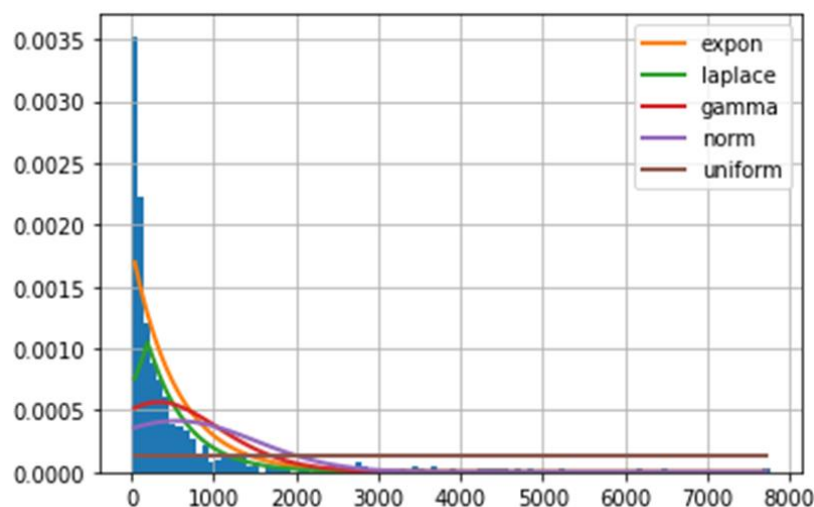
df = pd.read_excel('data.xlsx', sheet_name=0)
saoi = df ["peak social aoi"]

f = Fitter(saoi,
           distributions=['expon', 'gamma', 'norm', 'uniform', 'laplace'])
f.fit()
f.summary()

```

Κώδικας 7.1 – Υλοποίηση data fitting με Fitter.

Από τον Κώδικα 7.1, αρχικά προέκυψε το πως ταιριάζουν γραφικά οι κατανομές στο ιστόγραμμα των δεδομένων μας. Οπτικά, από το Σχήμα 7.8 βγαίνει το συμπέρασμα πως η εκθετική κατανομή είναι πιο κοντά στον τρόπο με τον οποίο κατανέμονται οι τιμές της P.



Σχήμα 7.8 – Fitting των κατανομών στο ιστόγραμμα των τιμών της Peak Social Age of Information.

Αυτό φαίνεται και από τον Πίνακα 7.1, όπου παρουσιάζονται οι 5 γνωστές και συνηθείς κατανομές που θέσαμε να κάνουν fit στα δεδομένα μαζί με τις τιμές του SSE για κάθε μια. Η εκθετική κατανομή έχει τη μικρότερη SSE και μάλιστα αρκετά ικανοποιητική για να προσδιορίσει τη διασπορά των δεδομένων.

distribution	sumsquare_error
exponential	0.000005
laplace	0.000010
gamma	0.000013
normal	0.000015
uniform	0.000020

Πίνακας 7.1 – Οι κατανομές με τις αντίστοιχες τιμές των SSE.

7.3.2 Distfit

Για να επιβεβαιωθεί η ορθότητα της βέλτιστης συνεχούς κατανομής, καθώς θα αποτελέσει βασικό χαρακτηριστικό για την μετέπειτα μελέτη, χρησιμοποιούμε ένα άλλο πακέτο της python, το distfit. Παρόμοια με το Fitter, κάνει fit τις κατανομές στα δεδομένα, βρίσκοντας την κατάλληλη. Για το peak SAoI, αυτό υλοποιείται ως εξής:

```

from distfit import distfit
import pandas as pd
import numpy as np

df = pd.read_excel('data.xlsx', sheet_name=0)
saoi = df ["social AoI"]

dist= distfit(distr=['expon', 'gamma', 'norm', 'uniform', 'laplace'])

dist.fit_transform(saoi) # Fit distributions on empirical data
dist.plot() # Plot the best fitted distribution

# All scores of the tested distributions
print(dist.summary)

# Distribution parameters for best fit
dist.model

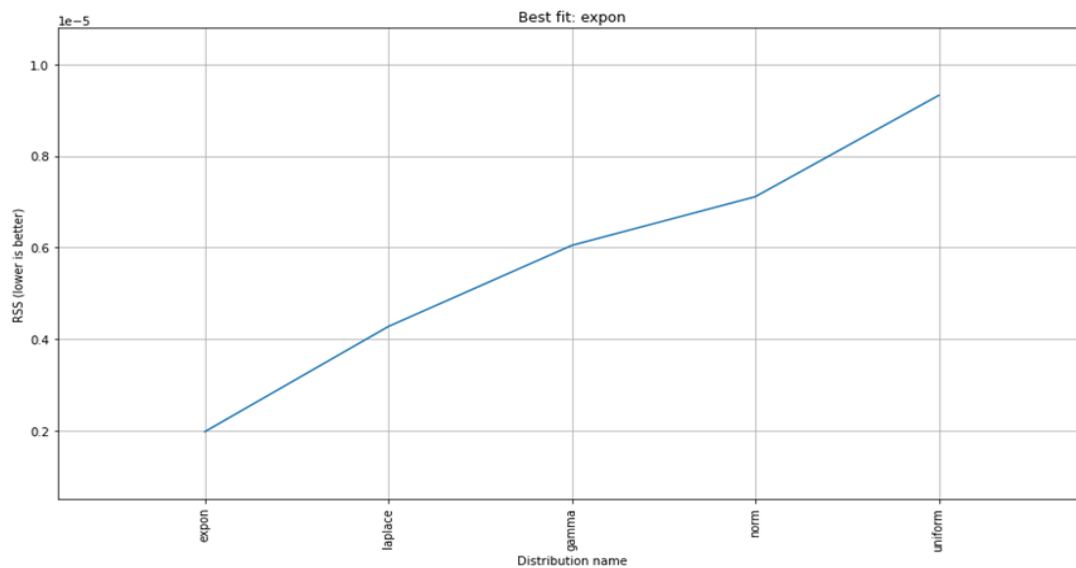
# Make plot
dist.plot_summary()

```

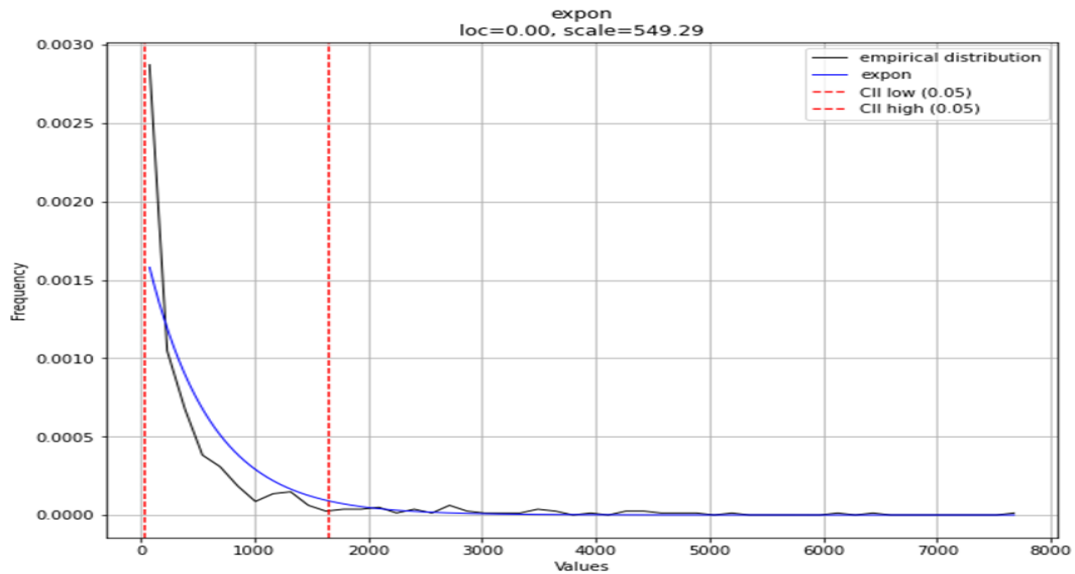
Κώδικας 7.2 – Υλοποίηση data fitting με distfit.

Η επιλογή της καλύτερης κατανομής γίνεται με βάση το RSS (Residual Sum of Square). Αυτό μετράει τη διαφορά μεταξύ των τιμών των δεδομένων και εκείνων που προκύπτουν από τις κατανομές. Ουσιαστικά είναι η ίδια μετρική με τη SSE, αφού τα errors της αναπαριστούν τα residuals («υπολλείματα») της RSS.

Στο Σχήμα 7.8 (α), βλέπουμε με αύξουσα σειρά του RSS τις κατανομές καθώς και τη γραμμή τάσης με την οποία αυξάνεται. Επαληθεύεται, και με αυτόν τον τρόπο ότι καλύτερο fit γίνεται με την εκθετική κατανομή, το οποίο φαίνεται γραφικά στο Σχήμα 7.8 (β), αφού έχει τη χαμηλότερη RSS.



(α)



(β)

Σχήμα 7.9 – Αποτελέσματα του distfit με:

(α) Απεικόνιση καλύτερων κατανομών με βάση το RSS

(β) Γραφικό fit της εκθετικής κατανομής στα δεδομένα

Με το distfit, ωστόσο, μπορούμε να αντλήσουμε και τις παραμέτρους των κατανομών ώστε να έχουμε όσο καλύτερο ταίριασμα γίνεται. Στον Πίνακα 7.2, εκτός από τα score (τιμές RSS) κάθε κατανομής, βλέπουμε και τις τιμές των location και scale που προσδιορίζουν σαφέστερα τα distributions.

Για την εκθετική κατανομή βλέπουμε το χαμηλότερο σκορ RSS και ίσο με $2 * 10^{-6}$ που υποδεικνύει αρκετά καλό data fitting. Βέβαια για να γίνει αυτό πρέπει οι παράμετροι scale και location να είναι κατάλληλες και συγκεκριμένα για την εκθετική βλέπουμε ότι προκύπτει: loc = 0 και scale = 549.292543.

distribution	score	loc	scale
exponential	0.000002	0.0	549.292543
laplace	0.000004	198.0	472.898662
gamma	0.000006	-2542.584179	170.829915
normal	0.000007	549.292543	963.549434
uniform	0.000009	0.0	7754.0

Πίνακας 7.2 – Οι κατανομές, οι RSS και οι παράμετροι τους.

Γενικά, ο τύπος για τη συνάρτηση πυκνότητας πιθανότητας (pdf) είναι:

$$f(x) = \frac{1}{scale} e^{-\frac{(x-loc)}{scale}}, x \geq loc \text{ και } scale > 0. \quad (7.1)$$

Το *loc* (=location) είναι μια παράμετρος που απλά μεταβάλλει το γράφημα της συνάρτησης δεξιά ή αριστερά στον οριζόντιο άξονα. Το *scale* είναι ίσο με: $scale = \frac{1}{lambda}$ και δείχνει το μέσο των δειγμάτων δηλαδή το χρόνο μεταξύ των διαδοχικών comments. Το *lambda* (=λ) ορίζει ρυθμό αφίξεων ανά μονάδα χρόνου, δηλαδή τα comments ανά δευτερόλεπτο. Συγκεκριμένα, έχουμε *scale* = 549.292543 sec μέσο χρόνο μεταξύ comments.

Επομένως, η pdf της εκθετικής κατανομής που κάνει fit στην peak Social Age of Information είναι:

$$f(x) = \frac{1}{549.292543} e^{-\frac{x}{549.292543}} \quad (7.2)$$

Κεφάλαιο 8: Value of Refresh

8.1 Μοντέλο για τη Value of Refresh

Ορίσαμε το Social Age Of Information ως το χρόνο που έχει περάσει από το τελευταίο comment που έχει γίνει στο post. Είδαμε ότι τα χρονικά διαστήματα που μεσολαβούν μεταξύ των comments μπορούμε να τα θεωρήσουμε interarrival times, των οποίων οι τιμές είναι αυτές του peak Social Age Of Information.

Ωστόσο, ένας χρήστης λαμβάνει την πληροφορία του comment τη στιγμή που έχει πρόσβαση στο post, αφού δεν είναι μόνιμα μέσα σε αυτό. Αυτό μας οδηγεί στο να δούμε τι πληροφορία παίρνει ένας χρήστης από το post που εξετάζουμε, σε σχέση με το ρυθμό που το «ανανεώνει».

Έστω Y η μεταβλητή που περιγράφει τους interarrival times μεταξύ των σχολίων. Είδαμε ότι οι χρόνοι αυτοί είναι εκθετικά κατανομημένοι με μέσο $E[Y] = 549.292543$ sec. Αυτό σημαίνει ότι η παραγωγή του περιεχομένου, δηλαδή ο αριθμός των νέων comments (διακριτή μεταβλητή) στο post σε δοσμένο χρονικό διάστημα Δt , ακολουθεί την κατανομή Poisson με ρυθμό $\lambda = 1 / E[Y] = 0.001820523531119555$ comments/sec.

Έτσι, ορίζουμε ως B τη μεταβλητή που περιγράφει τον αριθμό των comments του post σε χρονικό διάστημα T , το οποίο και θέτουμε να είναι ίσο με το διάστημα μεταξύ των refreshes. Με βάση τα όσα είπαμε παραπάνω, θα ακολουθεί κατανομή Poisson με ρυθμό $\lambda = 0.001820523531119555$ comments/sec.

Γνωρίζουμε πως το Facebook στα posts δεν αφήνει να δούμε απεριόριστο αριθμό νέων comments αφού αν είναι πάνω από έναν αριθμό M , πρέπει να κάνουμε εκ νέου ένα είδους refresh του post για να τα δούμε όλα τα καινούργια.

Ωστόσο, για να γίνει δυνατή η ανάλυση πρέπει να κάνουμε κάποιες υποθέσεις που θα συγκεκριμενοποιήσουν τη μελέτη:

- (1) Στο post βλέπουμε όσο το δυνατόν περισσότερο νέο περιεχόμενο. Δηλαδή, αν γίνονται νέα comments, ο χρήστης τα βλέπει μετά το refresh του, εκτός κι αν ξεπερνούν τον μέγιστο αριθμό M .

(2) Τα χρονικά διαστήματα, που περιγράφονται από τη μεταβλητή T , μεταξύ των refreshes είναι εκθετικά κατανομημένα με μέσο $1/\mu$.

(3) Τα παλαιά comments, όπως και συμβαίνει συνήθως στο Facebook, συμπληρώνουν την ποσότητα των συνολικών comments που εμφανίζονται αν δεν επαρκούν τα καινούργια.

Έτσι, μπορούμε να εισάγουμε μια νέα έννοια η οποία υποδεικνύει την αξία του refresh σε ένα post, σε σχέση με τις νέες πληροφορίες που θα πάρουμε σχετικά με αυτό, δηλαδή τα νέα comments. Το Value of Refresh (VoR), δηλαδή η αξία του refresh, είναι μία μετρική που της λέει πόσες νέες δραστηριότητες/αντιδράσεις βλέπουμε όταν κάνουμε ανανέωση της σελίδας. Ο αριθμός αυτός, λοιπόν, είναι της δείκτης για την ποσότητα χρήσιμης πληροφορίας που λαμβάνουμε κατά ένα refresh και θα το συμβολίζουμε και ως V . Η συγκεκριμένη μετρική αποτελεί συμβολή της παρούσας εργασίας.

Η ιδέα της αξίας μίας πράξης (όπως είναι το refresh), ως προς την απόκτηση νέας πληροφορίας, δεν είναι καινούργια αφού εισάγεται και στο [19] μία ανάλογη έννοια. Αυτή ήταν η αξία της πρόσβασης VoA (Value of Access) και γίνεται ανάλογη μελέτη στο πεδίο του timeline του Facebook.

Εκ πρώτου στοχασμού, φαίνεται λογικό η τιμή της VoR να εξαρτάται έμμεσα ή άμεσα από τη μετρική αυτή που ονομάσαμε Social Age of Information, κάτι το οποίο θα εξετάσουμε και κατά την ανάλυση της.

Στον παρακάτω πίνακα (Πίνακας 8.1) παρουσιάζονται οι μεταβλητές, με τις αντίστοιχες περιγραφές τους, που θα χρησιμοποιηθούν για την υλοποίηση του μοντέλου.

Μεταβλητή	Περιγραφή
V	ο αριθμός των νέων comments που βλέπει ο χρήστης ανά refresh
λ	ρυθμός δημοσίευσης comment
μ	ρυθμός του refresh
T	χρονικό διάστημα μεταξύ των refreshes
M	μέγιστος αριθμός προβολής comments ανά refresh
c	κόστος ανά refresh
B	αριθμός comments που δημιουργήθηκαν σε διάστημα T
ρ	μέση ποσότητα νέων comments ανά refresh

Πίνακας 8.1 – Οι μεταβλητές και η περιγραφή τους.

8.2 Υπολογισμός VoR

Για τον υπολογισμό της VoR, πρέπει να θεωρήσουμε σταθερό χρόνο μεταξύ των refreshes, τον οποίο και θέτουμε ίσο με τ ($T=\tau$). Θεωρούμε ότι σε αυτό το χρόνο δημιουργούνται $B(\tau) = B$ με δεδομένο μέγιστο αριθμό προβολής comments M .

Έχοντας ορίσει τη Value of Refresh (VoR), η τιμή που παίρνει ανά refresh είναι:

$$V = \begin{cases} B, & \text{αν } B < M \\ M, & \text{αλλιώς} \end{cases} \quad (8.1)$$

Όπως αποδείξαμε παραπάνω, η B ακολουθεί κατανομή Poisson με ρυθμό λ , επομένως:

$$P(B = b|\tau) = \frac{e^{-\lambda\tau} (\lambda\tau)^b}{b!} \quad (8.2)$$

Έτσι, για έναν τυχαίο αριθμό i , ισχύει:

$$P(V = i|\tau) = \begin{cases} P(B < i|\tau), & \text{αν } i < M \\ P(B \geq i|\tau), & \text{αν } i = M \\ 0 & , \text{αν } i > M \end{cases} \quad (8.3)$$

Από την παραπάνω σχέση, προκύπτει:

$$E(V|\tau) = (\sum_{i=0}^M iP(A = i|\tau)) + M P(B > M|\tau) \quad (8.4)$$

Όμως, ισχύει:

$$P(B > M|\tau) = 1 - P(B \leq M|\tau) \quad (8.5)$$

Επομένως:

$$E(V|\tau) = \left(\sum_{i=0}^M i \frac{e^{-\lambda\tau} (\lambda\tau)^i}{i!}\right) + M \left(1 - \sum_{i=0}^M i \frac{e^{-\lambda\tau} (\lambda\tau)^i}{i!}\right)$$

$$\Rightarrow E(V|\tau) = M - \sum_{i=0}^M (M-i) \frac{e^{-\lambda\tau} (\lambda\tau)^i}{i!} \quad (8.6)$$

Η $E(V|\tau)$ είναι η αναμενόμενη τιμή της V δεδομένου του χρόνου τ που ορίσαμε μεταξύ των refreshes, οπότε τη συμβολίζουμε ως \bar{V} .

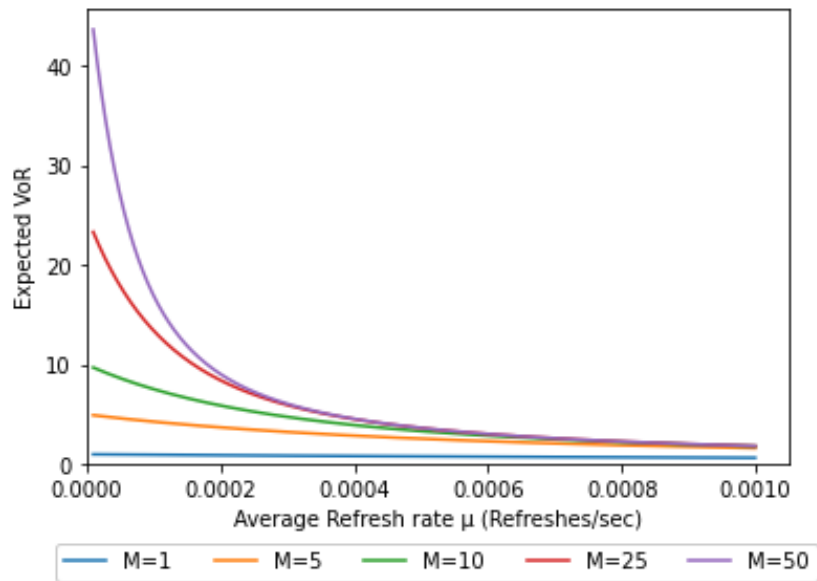
Από το νόμο ολικής πιθανότητας προκύπτει:

$$\bar{V} = \int_{\tau=0}^{\infty} E(V|\tau)f(\tau)d\tau \quad (8.7)$$

όπου $f(\tau) = \mu e^{-\mu\tau}$, η συνάρτηση πυκνότητας πιθανότητας για σταθερό χρόνο τ μεταξύ των refreshes, αφού η μεταβλητή T που τον περιγράφει, όπως θεωρήσαμε, ακολουθεί εκθετική κατανομή.

Αντικαθιστώντας τις τιμές και με τη βοήθεια του Παραρτήματος Β (Appendix Β) του [19], υπολογίζεται η τιμή του \bar{V} , ως:

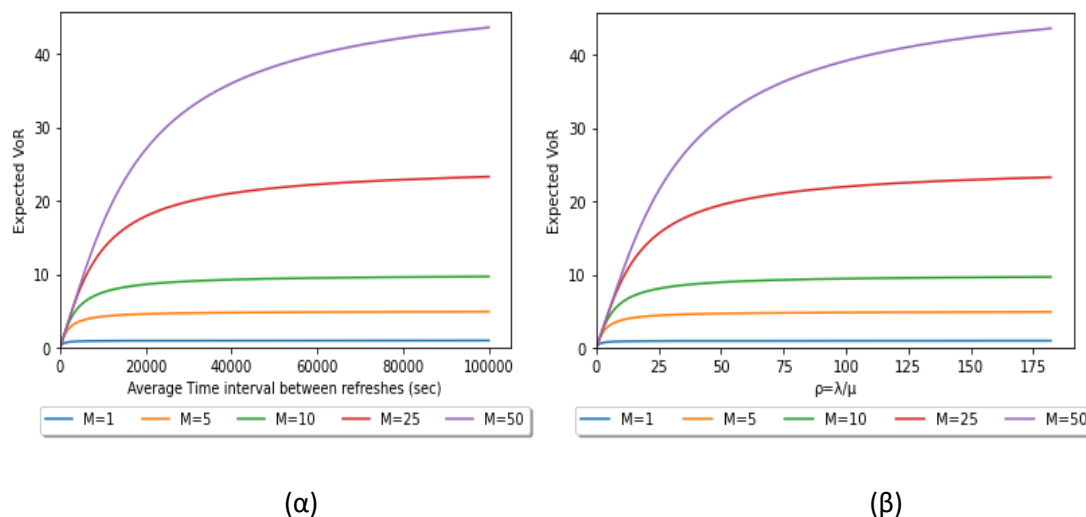
$$\bar{V} = \frac{\lambda}{\mu} \left(1 - \left(\frac{\lambda}{\lambda + \mu} \right)^M \right) \quad (8.8)$$



Σχήμα 8.1 – Η αναμενόμενη τιμή του VoR (\bar{V}) ως προς το μέσο ρυθμό μ των refreshes για $M = 1, 5, 10, 25, 50$.

Στο Σχήμα 8.1, βλέπουμε τις τιμές που παίρνει η μεταβλητή \bar{V} , αλλάζοντας τις τιμές των M, μ . Παρατηρούμε πως για πολύ χαμηλές τιμές του μ , δηλαδή όταν τα refreshes γίνονται πιο σπάνια, οι τιμές της αναμενόμενης αξίας του refresh τείνουν προς τον εκάστοτε μέγιστο αριθμό προβολής comments M . Αυτό είναι λογικό καθώς όσο μεγαλύτερο είναι το χρονικό διάστημα που μεσολαβεί μεταξύ των refreshes, θα έχουν δημοσιευτεί περισσότερα νέα comments με αποτέλεσμα ο χρήστης να παίρνει την υψηλότερη ποσότητα πληροφορίας. Αντίστοιχα, καθώς αυξάνονται οι τιμές του μ , έχουμε πιο συχνά refreshes, με αποτέλεσμα να εντοπίζονται όλο και λιγότερα νέα σχόλια στο καθένα. Έτσι, όπως φαίνεται από το παραπάνω σχήμα, μετά από

ορισμένες τιμές του μ , οι καμπύλες του VoR αρχίζουν να ταυτίζονται και τείνουν προς την ίδια χαμηλή τιμή (≤ 4), ανεξάρτητα από το M . Αυτό συμβαίνει γιατί για πιο υψηλό ρυθμό refresh, δεν μας περιορίζει ο μέγιστος αριθμός προβολής M , καθώς δεν είναι πολλά τα νέα comments.



Σχήμα 8.2 – (α) Η αναμενόμενη τιμή του VoR (\bar{V}) ως προς το μέσο χρονικό διάστημα $1/\mu$ μεταξύ των refreshes για $M = 1, 5, 10, 25, 50$.

(β) Η αναμενόμενη τιμή του VoR (\bar{V}) ως προς τη μέση ποσότητα ρ νέων comments ανά refresh για $M = 1, 5, 10, 25, 50$.

Ότι αναφέραμε παραπάνω, γίνεται αντιληπτό και από το Σχήμα 8.2, όπου φαίνονται οι τιμές της αναμενόμενης VoR (α) ως προς το μέσο χρονικό διάστημα μεταξύ των refreshes και (β) ως προς τη μέση ποσότητα ρ νέων comments ανά refresh.

8.3 Η Χρησιμότητα Περί Χρόνου

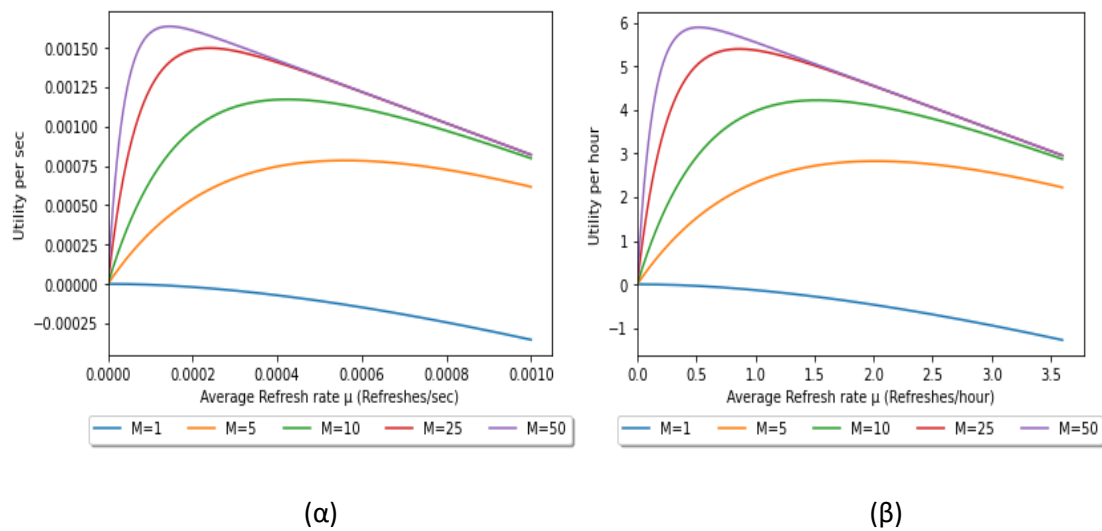
Η πραγματοποίηση ενός Refresh, ωστόσο, απαιτεί ορισμένες προϋποθέσεις. Αυτές μπορεί να είναι η χρήση δεδομένων κινητού δικτύου ή Wi-Fi, η κατανάλωση μπαταρίας της συσκευής καθώς ακόμα και η δαπάνη χρόνου από έναν χρήστη. Όλα αυτά τα «κόστη» τα συνοψίζουμε σε μια μεταβλητή, έστω c . Έτσι, το συνολικό κόστος c προκύπτει από το κάθε refresh ξεχωριστά και μετριέται με τις μονάδες μέτρησης του VoR. Το κόστος αυτό θα θεωρηθεί σταθερό για κάθε refresh και ίσο με $c=1$.

Έτσι, η χρησιμότητα (utility) ενός refresh προκύπτει αν από το Value of Refresh αφαιρέσουμε το κόστος c του refresh. Εμείς, όμως, θα υπολογίσουμε τη χρησιμότητα ανά μονάδα χρόνου που προκύπτει από τη διαφορά του κόστους του refresh από τον αναμενόμενο αριθμό νέων comments ανά refresh, πολλαπλασιασμένη με το ρυθμό των refreshes. Δηλαδή:

$$U(\bar{V}, \mu) = \mu(\bar{V} - c) \quad (8.9)$$

Λόγω της Σχέσης (8.8), προκύπτει:

$$U(\mu) = \lambda \left(1 - \left(\frac{\lambda}{\lambda + \mu} \right)^M \right) - \mu c \quad (8.10)$$



Σχήμα 8.3 – (α) Η χρησιμότητα ανά δευτερόλεπτο ενός χρήστη ως προς το μέσο ρυθμό μ των refreshes για $M = 1, 5, 10, 25, 50$.

(β) Η χρησιμότητα ανά ώρα ενός χρήστη ως προς το μέσο ρυθμό μ των refreshes για $M = 1, 5, 10, 25, 50$.

Στο Σχήμα 8.3 (α), όπου βλέπουμε τη χρησιμότητα ανά δευτερόλεπτο ως προς το ρυθμό μ , παρατηρούμε μεγαλύτερες τιμές χρησιμότητας για μεγαλύτερο M . Αυτό συμβαίνει γιατί, όπως είδαμε, έχουμε πιο υψηλή αναμενόμενη VoR για μεγαλύτερα

M ενώ το κόστος παραμένει σταθερό και ανεξάρτητο του M . Ωστόσο, αυξάνοντας το ρυθμό μ των refreshes, μικραίνει η χρησιμότητα λόγω της μείωσης του \bar{V} . Βλέπουμε, επίσης, πως για $M=1$, η χρησιμότητα ανά δευτερόλεπτο παίρνει αρνητικές τιμές, κάτι που σημαίνει ότι το «κόστος» του refresh είναι μεγαλύτερο από την ποσότητα πληροφορίας που παίρνουμε, δεδομένου του c που θέσαμε. Το γεγονός αυτό δικαιολογείται αν λάβουμε υπ' όψιν πως με $M=1$, μπορούμε να δούμε το πολύ 1 νέο comment κάθε φορά που κάνουμε refresh, το οποίο αποδεικνύεται ακόμα πιο «κοστοβόρο» αν τα refreshes γίνονται συχνότερα, αφού καθένα από αυτά έχει αρνητική επίπτωση στη χρησιμότητα. Στο Σχήμα 8.3 (β), παρουσιάζεται τη χρησιμότητα ανά ώρα ως προς το ρυθμό μ (επίσης σε refreshes/ώρα), ώστε να γίνει μια πιο εμφανής ποσοτικοποίηση τόσο της χρησιμότητας όσο και του μέσου ρυθμού refresh.

Όπως αντιλαμβανόμαστε, είναι σημαντικό η χρησιμότητα ανά μονάδα χρόνου να είναι όσο μεγαλύτερη γίνεται. Εφόσον παίρνουμε δεδομένες τιμές για το M και γνωρίζουμε τα λ, c , το $U(\mu)$ μεγιστοποιείται για κατάλληλη τιμή του μ .

Το $U(\mu)$ γίνεται μέγιστο όταν:

$$\frac{\partial U}{\partial \mu} = 0 \quad (8.11)$$

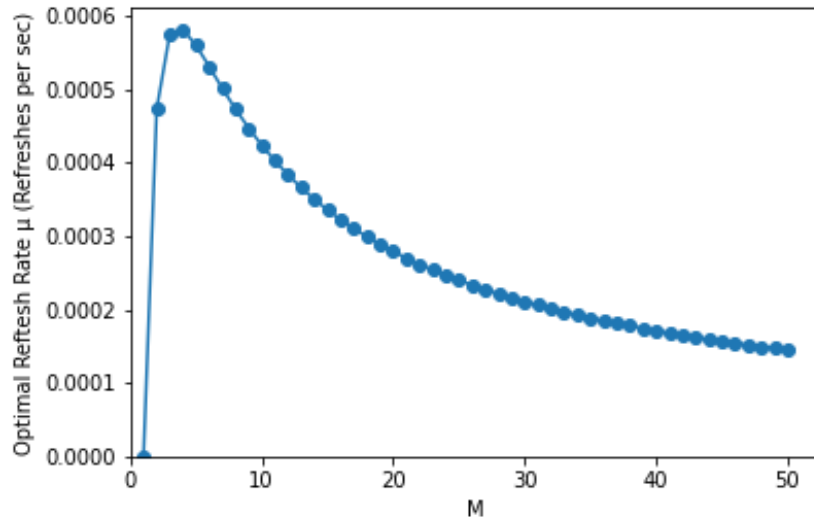
Λόγω της Σχέσης (8.10), έχουμε:

$$M \left(\frac{\lambda}{\lambda + \mu} \right)^{M+1} - c = 0 \quad (8.12)$$

Έτσι, η βέλτιστη τιμή του μ , δηλαδή εκείνη για την οποία μεγιστοποιείται η χρησιμότητα ανά μονάδα χρόνου, είναι:

$$\mu^* = \lambda \left(\frac{M+1 \sqrt{M}}{c+1 \sqrt{c}} - 1 \right) \quad (8.13)$$

Εφόσον γνωρίζουμε για το post, το ρυθμό δημοσίευσης των comments λ και έχοντας δεδομένο c , βρίσκουμε τη βέλτιστη τιμή του μ συναρτήσει του μέγιστου αριθμού προβολής comments ανά refresh (M).



Σχήμα 8.4 – Ο βέλτιστος μέσος ρυθμός των refreshes μ^* για τις διάφορες τιμές του μέγιστου αριθμού προβολής comments ανά refresh ($1 \leq M \leq 50$).

Από το Σχήμα 8.4, μπορούμε να δούμε πως για $M=1$, προκύπτει $\mu^*=0$ αφού, όπως αναλύσαμε προηγουμένως, η χρησιμότητα για αυτήν την τιμή είναι αρνητική. Παρατηρούμε πως έχουμε μια ανοδική πορεία για το βέλτιστο ρυθμό μ^* μέχρι την τιμή $M=4$, επειδή το μικρό σύνολο των σχολίων που μπορούμε να δούμε σε κάθε refresh είναι πιθανότερο να αποτελείται από νέα comments. Για $M \geq 5$, αν κάνουμε συχνότερα refresh αυξάνεται η πιθανότητα να βλέπουμε λιγότερα νέα comments ενώ καθώς το M μεγαλώνει θα εμφανίζονται παλαιά comments που δεν προσφέρουν χρήσιμη πληροφορία. Αυτός είναι ο λόγος που για αυτές τις τιμές ($M \geq 5$), παρατηρούμε μια συνεχής μείωση του μ^* . Έτσι, σε κάθε refresh για μεγάλο M , ο χρήστης θα αντλεί πολλά νέα comments, με αποτέλεσμα να μη χρειάζεται να βλέπει στο post τόσο συχνά.

8.4 Σχέση Peak Social Age of Information με Value of Refresh

Όπως είδαμε, η αξία του refresh VoR στο post που έχουμε, εξαρτάται από πολλούς παράγοντες, οι οποίοι καθορίζουν την τιμή του με διαφορετικούς τρόπους. Εμείς έως τώρα, γνωρίζοντας το μέσο ρυθμό δημοσίευσης των comments λ στο post, αρκεστήκαμε να δούμε τη συμπεριφορά του \bar{V} , συναρτήσεϊ άλλων μεταβλητών.

Όμως, η τιμή αυτή του λ , έχει άμεση σχέση με τις τιμές του peak Social Age of Information (peak SAoI), κάτι που οδηγεί στην αναζήτηση σύνδεσης του με τη \bar{V} .

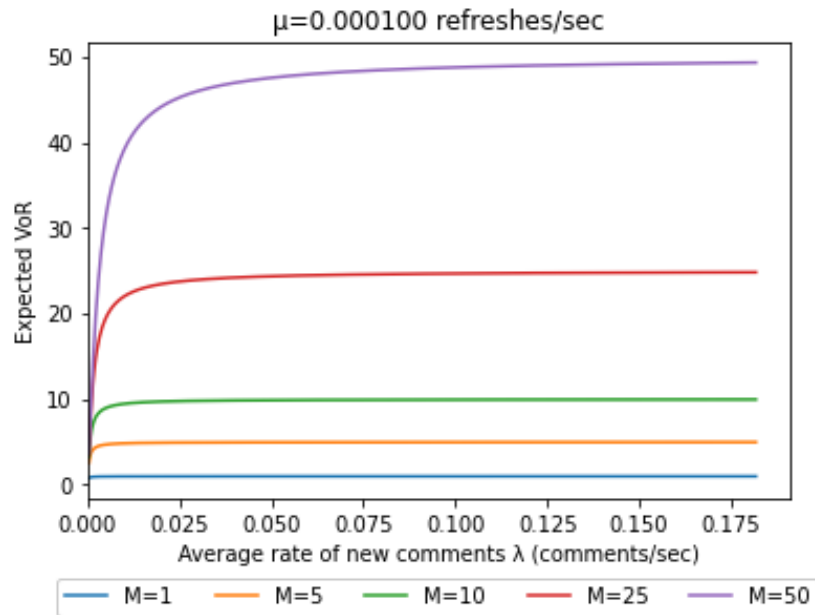
Η μεταβλητή Y που περιγράφει τους interarrival times μεταξύ των σχολίων, όπως είπαμε περιγράφει τις τιμές του peak SAoI. Αυτοί οι χρόνοι είναι εκθετικά καταναμημένοι με μέσο $E[Y] = \frac{1}{\lambda} = 549.292543$ που είναι ο μέσος όρος των τιμών του peak SAoI.

Έτσι, έχοντας ως P τη μεταβλητή που περιγράφει τις τιμές του peak SAoI, έχουμε ότι η μέση τιμή της, δηλαδή η average peak Social Age of Information είναι:

$$\bar{P} = \frac{1}{\lambda} \quad (8.14)$$

Δηλαδή, βάσει του post που είναι στη διάθεση μας, έχουμε $\bar{P} = 549.292543$ sec.

Πρώτα απ' όλα, θα δούμε πως μεταβάλλεται η \bar{V} , σε συνάρτηση με το μέσο ρυθμό δημοσίευσης των comments λ , για μια δεδομένη τιμή του μ .

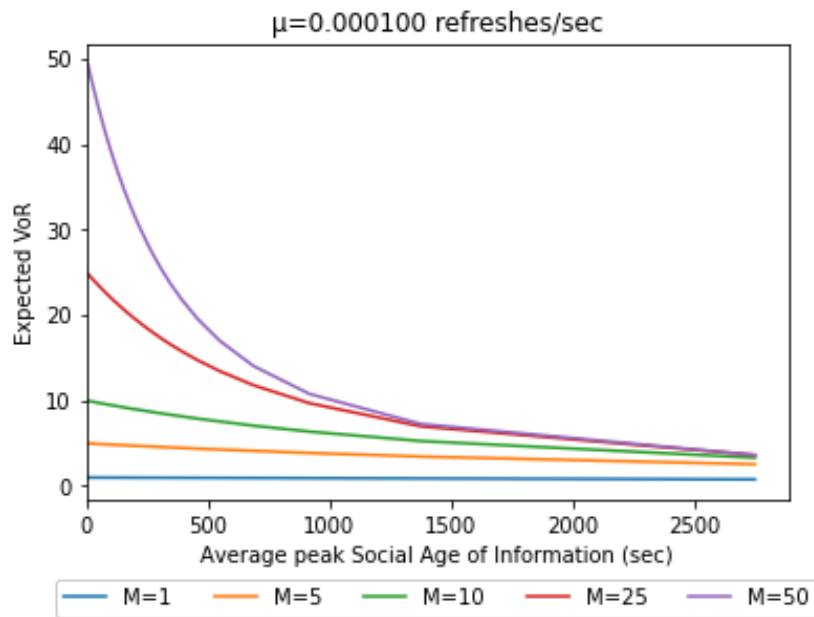


Σχήμα 8.5 – Η αναμενόμενη τιμή του VoR (\bar{V}) ως προς το μέσο ρυθμό λ δημοσίευσης comments για $\mu=0.0001$ και $M = 1, 5, 10, 25, 50$.

Στο παραπάνω σχήμα (Σχήμα 8.5), ξεφεύγουμε από τα δεδομένα του post που εξετάστηκε, δοκιμάζοντας διαφορετικές τιμές του λ . Είναι εμφανές πως αυξάνοντας το ρυθμό δημοσίευσης των comments επιτυγχάνονται καλύτερες τιμές για την αναμενόμενη VoR, καθώς εμφανίζονται όλο και περισσότερα νέα comments. Τότε μάλιστα, όπως είναι λογικό, η \bar{V} τείνει προς το εκάστοτε M , αφού το πλήθος των νέων comments πλησιάζει ή και ξεπερνά τον μέγιστο αριθμό προβολής comments ανά refresh.

Εισάγοντας στην (8.8), την (8.14) $\Rightarrow \lambda = \frac{1}{\bar{P}}$, προκύπτει:

$$\bar{V} = \frac{\lambda}{\mu \bar{P}} \left(1 - \left(\frac{1}{1 + \mu \bar{P}} \right)^M \right) \quad (8.15)$$



Σχήμα 8.6 – Η αναμενόμενη τιμή του VoR (\bar{V}) ως προς τη μέση peak Social Age of Information για $\mu=0.0001$ και $M = 1, 5, 10, 25, 50$.

Γενικότερα η μέση peak Social AoI είναι σημαντικό να διατηρεί χαμηλές τιμές ώστε να υπάρχουν «φρέσκιες» πληροφορίες. Αυτό επιβεβαιώνεται και στην περίπτωση της αναμενόμενης αξίας του refresh, η οποία αυξάνεται όσο πιο μικρή είναι η τιμή της \bar{P} . Το γεγονός αυτό το βλέπουμε στο Σχήμα 8.6, όπου η αναμενόμενη VoR ανά refresh παίρνει μέγιστη τιμή $\bar{V}_{max} \rightarrow M$ καθώς $\bar{P} \rightarrow 0$. Στη συνέχεια και καθώς αυξάνεται η

μέση peak SAoI, τα διαστήματα μεταξύ των comments είναι μεγαλύτερα, οπότε εντοπίζεται και χαμηλότερη \bar{V} .

Αυτό σημαίνει πως στο post που εξετάζουμε, αν η μέση peak SAoI ήταν μικρότερη, δηλαδή αν είχαμε μικρότερα διαστήματα μεταξύ των σχολίων ($\bar{P} < 549.292543$ sec), ένας χρήστης θα έβλεπε περισσότερα νέα comments σε κάθε refresh του ($\bar{V} \uparrow$).

Μη ξεχνάμε, βέβαια, πως για να ισχύει το μοντέλο, και ως εκ τούτου οι σχέσεις που έχουν αποδειχθεί, πρέπει οι τιμές του peak Social Age of Information να ακολουθούν εκθετική κατανομή.

Η πορεία που ακολουθεί η \bar{V} ως προς την \bar{P} , επιβεβαιώνεται και στον Πίνακα 8.2, βλέποντας τις ακριβείς τιμές, όπως προκύπτουν από τη Σχέση (8.15) για $M=25$. Με μπλε χρώμα παρουσιάζονται οι τιμές που έχουμε στο post που αναλύουμε.

$M=25$

Average Peak Social Age of Information (sec)	Expected Value of Refresh (\bar{V})
1	24.967529229541483
100	22.023155700621434
500	14.093944566044748
549.292543	13.423198683122536
1000	9.077040018229358
1500	6.464149085270139
2000	4.94758701994802
2500	3.984888427254817
3000	3.3286094959623322

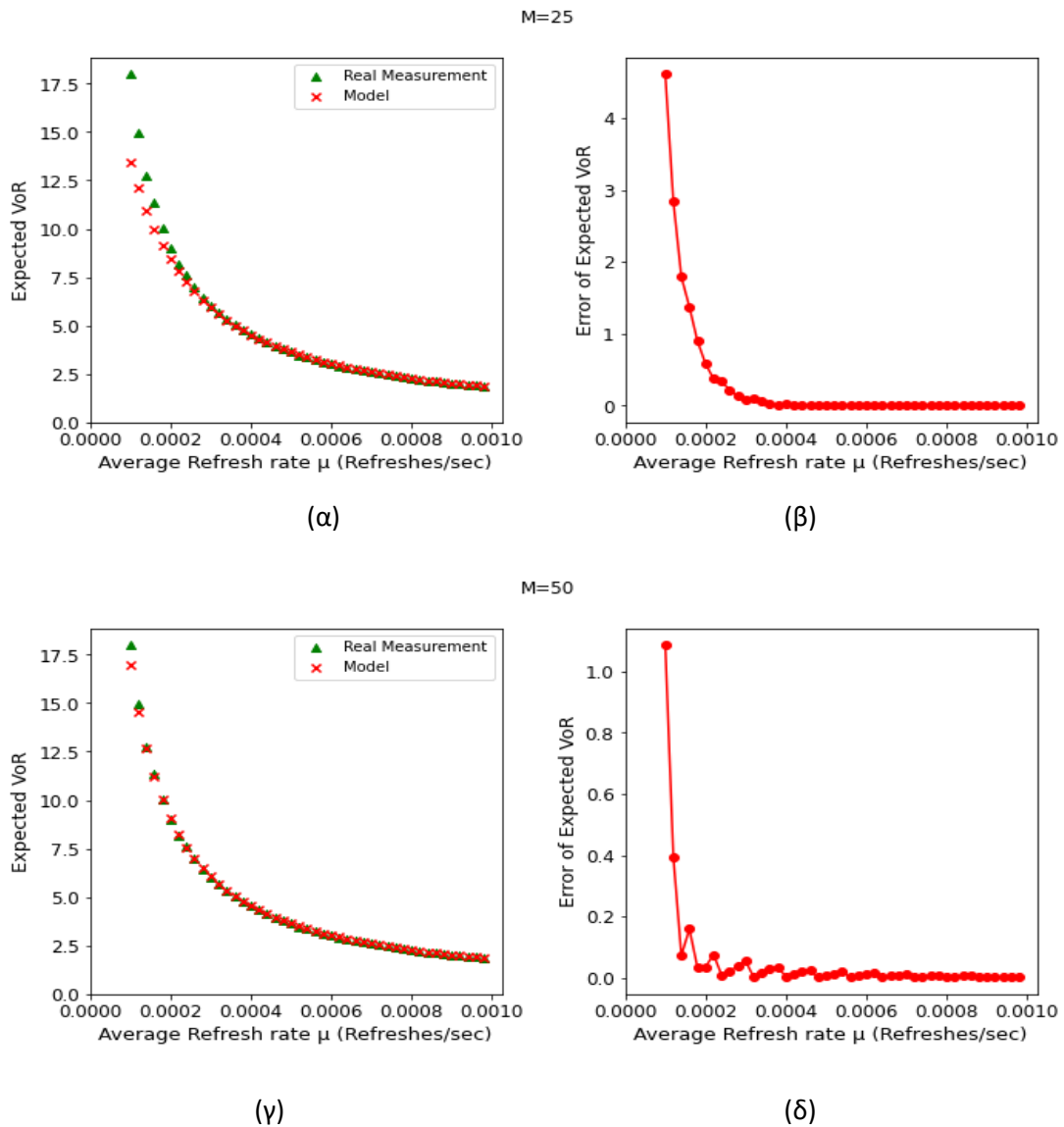
Πίνακας 8.2 – Οι αναμενόμενες τιμές του VoR για διάφορες τιμές Average Peak Social Age of Information για $M=25$.

8.5 Προσομοίωση και Σύγκριση με το Μοντέλο

Η μετρική Value of Refresh που χρησιμοποιήσαμε, ποσοτικοποίησε την αξία του refresh που κάνει ένας χρήστης, σχετικά με τις νέες πληροφορίες που παίρνει, για το post που εξετάζουμε. Ωστόσο, για τη μέτρηση του, χρησιμοποιήσαμε μόνο το μέσο ρυθμό λ δημοσίευσης νέων comments, όπως προκύπτει από το post. Γνωρίζοντας, όμως, τους χρόνους μεταξύ των comments (peak Social Age of Information), μπορούμε να μετρήσουμε με ακρίβεια την αξία καθενός από τα refresh ξεχωριστά. Να παρατηρήσουμε, δηλαδή, ακριβώς πόσα νέα σχόλια βλέπει ο χρήστης σε κάθε refresh του, παίρνοντας όμως σταθερά χρονικά διαστήματα μεταξύ τους και ίσα με το αντίστοιχο μέσο $\frac{1}{\mu}$. Η μέση τιμή αυτών των μετρήσεων, μπορεί να αποτελέσει μια καλή σύγκριση για τα αποτελέσματα της αναμενόμενης τιμής του VoR (\bar{V}) που εξήχθησαν από το μοντέλο νωρίτερα.

Δοκιμάζουμε διαφορετικές τιμές του M για διάφορα εύρη του ρυθμού μ των refreshes, ώστε να συγκρίνουμε πολύπλευρα τους δύο διαφορετικούς τρόπους μέτρησης της αναμενόμενης VoR. Στα γραφήματα της αναμενόμενης VoR, με κόκκινο χρώμα προβάλλονται οι τιμές που προήλθαν από το μοντέλο ενώ με πράσινο χρώμα εκείνες που προκύπτουν με τις μετρήσεις στα σταθερά χρονικά διαστήματα. Επίσης, απεικονίζεται γραφικά και η απόκλιση του μοντέλου από τις μετρήσεις μέσω του σφάλματος (error).

Αρχικά, εξετάζεται τι συμβαίνει για χαμηλότερες τιμές του ρυθμού μ των refreshes, δηλαδή για μεγαλύτερα χρονικά διαστήματα μεταξύ αυτών.



Σχήμα 8.7 – (α) Η αναμενόμενη τιμή του VoR ως προς το μέσο ρυθμό μ των refreshes για $0.0001 \leq \mu \leq 0.001$ και $M = 25$.

(β) Το σφάλμα της αναμενόμενης τιμής του VoR ως προς το μέσο ρυθμό μ των refreshes για $0.0001 \leq \mu \leq 0.001$ και $M = 25$.

(γ) Η αναμενόμενη τιμή του VoR ως προς το μέσο ρυθμό μ των refreshes για $0.0001 \leq \mu \leq 0.001$ και $M = 50$.

(δ) Το σφάλμα της αναμενόμενης τιμής του VoR ως προς το μέσο ρυθμό μ των refreshes για $0.0001 \leq \mu \leq 0.001$ και $M = 50$.

Στο Σχήμα 8.7 (α), παρατηρούμε τις τιμές του expected VoR για $M=25$ και βλέπουμε πως δεν ταυτίζονται πλήρως για τους δύο διαφορετικούς τρόπους μέτρησης. Πιο συγκεκριμένα, οι «πραγματικές» μετρήσεις μας δίνουν μεγαλύτερη \bar{V} , με διακριτές διαφορές για χαμηλότερο ρυθμό μ (≤ 0.00022). Ωστόσο, καθώς μεγαλώνει ο ρυθμός

μ των refreshes (μέχρι $\mu=0.001$), εντοπίζεται μια καλή ταύτιση μεταξύ των τιμών των δύο μετρήσεων. Αυτό, μπορεί να διακριθεί και από το Σχήμα 8.7 (β), όπου εμφανίζεται το error της αναμενόμενης VoR σε σχέση με το μ για $M=25$. Έτσι, για μικρό μ , το error παίρνει μεγαλύτερες τιμές που δεν μπορούν να αγνοηθούν ενώ με την αύξηση του μ , αρχίζει και κυμαίνεται γύρω από το μηδέν.

Στη συνέχεια, για $M=50$ στο Σχήμα 8.7 (γ), βλέπουμε ότι οι διαφορετικές μετρήσεις είναι ακόμα πιο κοντά και για τις μικρότερες τιμές του μ . Το τελευταίο αποτυπώνεται και στο Σχήμα 8.7 (δ), με το error να είναι σημαντικά μικρότερο καταλήγοντας φυσικά να είναι αμελητέο καθώς ο ρυθμός των refresh είναι μεγαλύτερος.

Παρακάτω παρουσιάζεται το μέσο σφάλμα για διάφορες τιμές του M και για δύο διαφορετικά διαστήματα τιμών του ρυθμού των refreshes μ .

$0.0001 \leq \mu \leq 0.001$

M	Average Error
10	1.0067857875741055
25	0.3030629798622683
50	0.05101488019145292
75	0.025645907433558643
87	0.02433528514558436
100	0.027341689288347737
125	0.029437144347962955

Πίνακας 8.3 – Το μέσο Error για διάφορες τιμές του M και $0.0001 \leq \mu \leq 0.001$.

Στον Πίνακα 8.3, για $0.0001 \leq \mu \leq 0.001$, δηλαδή για μέσο χρονικό διάστημα $1000 \text{ sec} \leq 1/\mu \leq 10000 \text{ sec}$ μεταξύ των refreshes, οι τιμές του average error μειώνεται καθώς ο μέγιστος αριθμός προβολής comments αυξάνεται μέχρι το $M=87$. Εκεί παρατηρείται και η μεγαλύτερη ταύτιση, για τις παραπάνω τιμές του μ . Για $M>87$, το μέσο error αυξάνεται ελάχιστα. Αυτό συμβαίνει, επειδή για μικρό M , το μοντέλο δίνει τιμές της \bar{V} μικρότερες από την πραγματική μέτρηση πάνω στο dataset, οι οποίες όμως τις προσεγγίζουν καθώς το M αυξάνεται και εν τέλει τις προσπερνούν. Έτσι στο

ενδιάμεσο ($M=87$) προκύπτει η καλύτερη συμφωνία αυτών των τιμών, άρα και το μικρότερο μέσο σφάλμα.

Αξίζει να δούμε τη σύγκριση για ακόμα μεγαλύτερο ρυθμό μ , η οποία προβάλλεται στον παρακάτω πίνακα.

$0.001 \leq \mu \leq 0.01$

M	Average Error
10	0.0010955922121458326
14	0.0004991348773580496
25	0.0006222637592744595
50	0.0006231064449487443
75	0.0006231064579007813
100	0.0006231064579010034
125	0.0006231064579010034

Πίνακας 8.4 – Το μέσο Error για διάφορες τιμές του M και $0.001 \leq \mu \leq 0.01$.

Στον Πίνακα 8.4, για $0.001 \leq \mu \leq 0.01$, δηλαδή για μέσο χρονικό διάστημα $100 \text{ sec} \leq 1/\mu \leq 1000 \text{ sec}$ μεταξύ των refreshes, οι τιμές του average error είναι αμελητέες για $M > 10$, δηλαδή τόσο το μοντέλο όσο και η προσομοίωση δίνουν σχεδόν ίδια αποτελέσματα. Η καλύτερη ταύτιση γίνεται για $M=14$.

Κεφάλαιο 9: Σύνοψη

9.1 Συμπεράσματα

Στην παρούσα διπλωματική εργασία μελετήθηκε η ηλικία της πληροφορίας προσαρμοσμένη στον τομέα των κοινωνικών δικτύων. Λαμβάνοντας έτοιμα δεδομένα για τα σχόλια μίας δημοσίευσης στο Facebook, εξετάστηκε η κοινωνική ηλικία της πληροφορίας, όπως την αντιλαμβάνεται ο χρήστης. Αναπτύχθηκε παράλληλα ένα μοντέλο για την εκτίμηση της αξίας της ανανέωσης (refresh) της δημοσίευσης, ως προς την απόκτηση νέων πληροφοριών. Αυτό έγινε δεδομένου της εύρεσης εκθετικής κατανομής για τη μέγιστη κοινωνική ηλικία της πληροφορίας της δημοσίευσης. Τα αποτελέσματα που βγήκαν από το μοντέλο συγκρίθηκαν με εκείνα που εξάγονται υπολογίζοντας ακριβώς την ποσότητα κάθε refresh ξεχωριστά.

Παρατηρήθηκε μια αρκετά καλή εφαρμογή του μοντέλου στο post. Δεδομένου βέβαια μερικών υποθέσεων, το μοντέλο αντλεί ρεαλιστική ποσοτικοποίηση της αξίας του refresh σχετικά με τα comments που γίνονται στο post, βγάζοντας σημαντικά συμπεράσματα τόσο για τη συχνότητα που είναι καλύτερο να γίνονται τα refreshes όσο και για τη χρησιμότητα τους. Είδαμε πως όσο μικρότερη είναι η μέγιστη κοινωνική ηλικία της πληροφορίας, τόσο πιο «φρέσκιες» πληροφορίες λαμβάνει ο χρήστης που βλέπει το post, ενώ έτσι η αξία της πρόσβασης σε αυτό μεγιστοποιείται. Έγινε φανερός ο τρόπος με τον οποίο, η πληροφορία που λαμβάνει ένας χρήστης καθώς και η στιγμή που αυτό γίνεται πράξη, εξαρτώνται από διάφορους παράγοντες, όπως ο ρυθμός πρόσβασης σε αυτά. Ο τελευταίος αποδείχθηκε πως δεν μπορεί να είναι εντελώς αυθαίρετος αφού λήφθηκε υπόψιν και ένα δεδομένο κόστος που υπάρχει από την ανανέωση μίας δημοσίευσης.

9.2 Μελλοντικές Επεκτάσεις

Στα πλαίσια της διπλωματικής εργασίας δόθηκε μια πιθανή χρήση της κοινωνικής ηλικίας της πληροφορίας και, τηρουμένων μερικών συνθηκών, η συσχέτιση της με την ποσοτικοποίηση της αξίας ενός refresh σε μια δημοσίευση. Αυτή η μελέτη μπορεί να βοηθήσει στο γενικότερο προσδιορισμό της βέλτιστης συχνότητας με την οποία ένας χρήστης θα ήταν καλό να επισκέπτεται τα κοινωνικά δίκτυα. Βέβαια, μπορεί να

δοθεί ακόμα μεγαλύτερη σημασία στον τρόπο και την ποσότητα με τα οποία εμφανίζονται στους χρήστες δημοσιεύσεις, σχόλια ή οποιαδήποτε άλλη ενέργεια, έτσι ώστε να εξαχθούν ακόμα πιο χρήσιμα αποτελέσματα. Να υπολογιστούν, δηλαδή, περισσότερα χαρακτηριστικά, βάσει των οποίων λειτουργούν τα κοινωνικά δίκτυα, τα οποία θα το κάνουν πιο ρεαλιστική και αλλά και πιο σύνθετη τη μελέτη. Συγκεκριμένα, στο Facebook, θα μπορούσε να ληφθεί υπόψιν ο αλγόριθμος ο οποίος χρησιμοποιείται για την εμφάνιση των comments σε ένα post.

Επίσης, ο ρυθμός πρόσβασης των χρηστών στα κοινωνικά δίκτυα έχει σημαντικές επιπτώσεις στην κίνηση των πακέτων που διέρχονται μέσα στο δίκτυο. Με αυτόν τον τρόπο δημιουργείται άμεση σχέση με την ηλικία της πληροφορίας που φτάνει στον προορισμό μέσω του δικτύου. Έτσι, η έρευνα της κοινωνικής ηλικίας της πληροφορίας, όπως εισηγήθηκε στην εργασία, και του ρυθμού ανανέωσης των σελίδων κοινωνικών δικτύων, μπορεί να επεκταθεί με σκοπό τη μείωση της συμφόρησης που υφίσταται το δίκτυο, άρα και της ηλικίας της πληροφορίας που μεταφέρεται σε εκείνο.

Βιβλιογραφία

- [1] S. Kaul, R. Yates and M. Gruteser, "Real-time status: How often should one update?," 2012 Proceedings IEEE INFOCOM, 2012, pp. 2731-2735, doi: 10.1109/INFOCOM.2012.6195689.
- [2] M. Costa, M. Codreanu and A. Ephremides, "On the Age of Information in Status Update Systems With Packet Management," in IEEE Transactions on Information Theory, vol. 62, no. 4, pp. 1897-1910, April 2016, doi: 10.1109/TIT.2016.2533395.
- [3] M. Costa, M. Codreanu and A. Ephremides, "Age of information with packet management," 2014 IEEE International Symposium on Information Theory, 2014, pp. 1583-1587, doi: 10.1109/ISIT.2014.6875100.
- [4] Nelson, R. 1995. Probability, stochastic processes and queueing theory: the mathematics of computer performance modeling. New York: Springer-Verlang
- [5] S. K. Kaul, R. D. Yates and M. Gruteser, "Status updates through queues," 2012 46th Annual Conference on Information Sciences and Systems (CISS), 2012, pp. 1-6, doi: 10.1109/CISS.2012.6310931.
- [6] K. Chen and L. Huang, "Age-of-information in the presence of error," 2016 IEEE International Symposium on Information Theory (ISIT), 2016, pp. 2579-2583, doi: 10.1109/ISIT.2016.7541765.
- [7] S. Kaul, M. Gruteser, V. Rai and J. Kenney, "Minimizing age of information in vehicular networks," 2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2011, pp. 350-358, doi: 10.1109/SAHCN.2011.5984917.
- [8] C. Kam, S. Kompella, G. D. Nguyen and A. Ephremides, "Effect of Message Transmission Path Diversity on Status Age," in IEEE Transactions on Information

Theory, vol. 62, no. 3, pp. 1360-1374, March 2016, doi: 10.1109/TIT.2015.2511791.

[9] Antzela Kosta; Nikolaos Pappas; Vangelis Angelakis, *Age of Information: A New Concept, Metric, and Tool*, now, 2017.

[10] Tripathi, Vishrant & Talak, Rajat & Modiano, Eytan. (2019). Age of Information for Discrete Time Queues.

[11] Talak, Rajat & Karaman, Sertac & Modiano, Eytan. (2018). Can Determinacy Minimize Age of Information?.

[12] R. Talak, S. Karaman and E. Modiano, "Optimizing Information Freshness in Wireless Networks Under General Interference Constraints," in *IEEE/ACM Transactions on Networking*, vol. 28, no. 1, pp. 15-28, Feb. 2020, doi: 10.1109/TNET.2019.2946481.

[13] Age of Information for Real-time Network Applications, Zhong, Jing. Rutgers The State University of New Jersey, School of Graduate Studies. ProQuest Dissertations Publishing, 2019. 22624908.

[14] I. Kadota, E. Uysal-Biyikoglu, R. Singh and E. Modiano, "Minimizing the Age of Information in broadcast wireless networks," 2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2016, pp. 844-851, doi: 10.1109/ALLERTON.2016.7852321.

[15] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal and N. B. Shroff, "Update or Wait: How to Keep Your Data Fresh," in *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7492-7508, Nov. 2017, doi: 10.1109/TIT.2017.2735804.

[16] B. Barakat, S. Keates, I. Wassell and K. Arshad, "Is the Zero-Wait Policy Always Optimum for Information Freshness (Peak Age) or Throughput?," in *IEEE Communications Letters*, vol. 23, no. 6, pp. 987-990, June 2019, doi: 10.1109/LCOMM.2019.2907935.

[17] Wikipedia, API, <https://en.wikipedia.org/wiki/API>

[18] Djordje Padejski. (2018).

Dataset: 24-09-to-10-11-2016-endingthefed_facebook_statuses

File: 24-09-to-10-11-2016-endingthefed_facebook_comments.csv

Ανακτήθηκε από: <https://data.world/padejski/24-09-to-10-11-2016-endingthefedfacebookstatuses>

[19] E. Hargreaves, D. S. Menasché and G. Neglia, "How Often Should I Access My Online Social Networks?," 2019 IEEE 27th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), 2019, pp. 189-200, doi: 10.1109/MASCOTS.2019.00028