M162 - Τεχνικές Ιδιωτικότητας 2019

Project, **Part A**

Κυριάκος Χριστοδούλου
cs2180019

Crowds Anonymous Communication Model Simulation

FBLEAU estimation of adversary’s success and Information Leakage

Πίνακας Περιεχομένων

Crowds

Το Crowds (Image 1) είναι ένα μοντέλο ανώνυμης επικοινωνίας που μπορεί να είναι αποτελεσματικό, ειδικά σε εφαρμογές ανώνυμης πλοήγησης στο διαδίκτυο, καθώς παρέχει κυρίως sender anonymity.

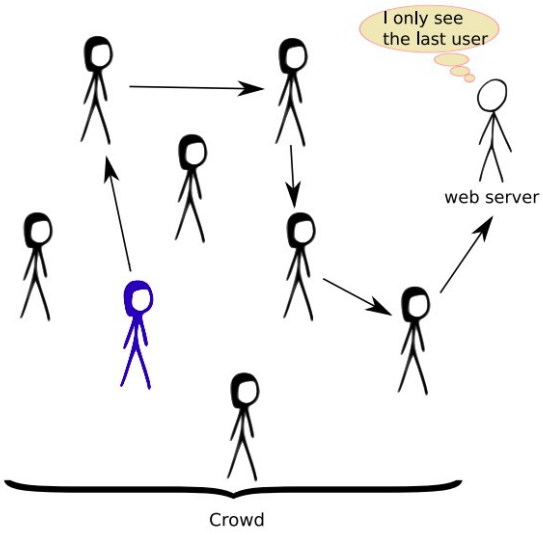


Image 1: Crowds

Ο τρόπος που λειτουργεί είναι αρκετά απλός:

1. Ο αποστολέας προωθεί το request, που θέλει να στείλει στον web server, σε κάποιο χρήστη του Crowds (συμπεριλαμβανομένου και του εαυτού του) με τον οποίο μπορεί να επικοινωνήσει (έστω την Alice)
2. Η Alice επιλέγει να προωθήσει το request σε κάποιο δικό της γείτονα (συμπεριλαμβανομένου και του εαυτού της) με πιθανότητα Φ ενώ με πιθανότητα 1-Φ το στέλνει στον web server (αντίπαλος). Αυτό επαναλαμβάνεται μέχρις ότου φτάσει τελικά το request στον web server
3. Ο αντίπαλος εντοπίζει σαν αποστολέα τον τελευταίο χρήστη

Ο αντίπαλος από την άλλη μπορεί να έχει υπό τον έλεγχό του και κάποιους από τους χρήστες του Crowds (corrupted). Με αυτό τον τρόπο εντοπίζει με μεγαλύτερη πιθανότητα τον αρχικό αποστολέα του μηνύματος.

Σκοπός του Crowds είναι ο αποστολέας “στα μάτια του αντιπάλου” να έχει τις ίδιες πιθανότητες να έχει στείλει το request όπως και οι υπόλοιποι χρήστες του Crowds (Probable innocence). Σε περίπτωση που κάποιος χρήστης γίνει detected από κάποιον corrupted user, καταστρέφεται το μονοπάτι (borken path) και επαναλαμβάνεται η ίδια διαδικασία, είτε από τον ίδιο (last honest) είτε από τον αρχικό αποστολέα (initiator).

Περιγραφή προσομοίωσης

Η προσομοίωση παίρνει τις εξής παραμέτρους όπως ζητείται στην εκφώνηση:

1. παράμετρος **Φ**, η πιθανότητα ένας χρήστης να προωθήσει το μήνυμα παρά να το στείλει στον τελικό παραλήπτη
2. ο **γράφος,** με τις ακμές μεταξύ των χρηστών
3. τα **id** των **corrupted** χρηστών
4. οι **senders** όλων των εκτελέσεων του πρωτοκόλλου
5. ο αριθμός των πιθανών broken-paths κάθε εκτέλεσης
6. **fix-strategy**, ο τρόπος με τον οποίο επιδιορθώνει ένας χρήστης ένα broken-path

Δημιουργώ μία λίστα με τους senders, και μία λίστα με τους corrupted χρήστες, καθώς και ένα dictionary για την αναπαράσταση του γράφου ως εξής: για τον χρήστη με id=0, αντιστοιχίζω το 0 με μία λίστα η οποία περιέχει τους γείτονες του 0 (πχ 0: [3,6,7]).

Ακολούθως παίρνω με τη σειρά, 1 προς 1 τους senders, και επιλέγω με τυχαίο τρόπο ένα από τους γείτονές του. Αν αυτός ανήκει στους corrupted χρήστες, τότε θεωρώ ότι ο sender έγινε **detected**.

Ανάλογα με την παράμετρο broken-paths, επαναλαμβάνεται η ίδια διαδικασία μέχρι να προωθήσει το μήνυμα σε μη-corrupted χρήστη ή μέχρι να αναλωθούν τα broken-paths. Αν καταφέρει να προωθήσει το μήνυμα, τότε με πιθανότητα **φ**, ο παραλήπτης επιλέγει ένα από τους γείτονές του ενώ με πιθανότητα **1-φ** παραδίδει το μήνυμα στον τελικό παραλήπτη (οπότε και γίνεται **detected**).

Αν επιλέξει να το προωθήσει, πάλι επιλέγει τυχαία ένα από τους γείτονές του και ακολουθείται η ίδια διαδικασία, με την διαφορά ότι στην περίπτωση detection, και εφόσον τα broken-paths είναι >0, ανάλογα με την παράμετρο **fix-strategy** επιλέγει είτε ο ίδιος ένα από τους γείτονές του (last_honest) είτε η διαδικασία ξεκινάει από τον αρχικό αποστολέα (initiator).

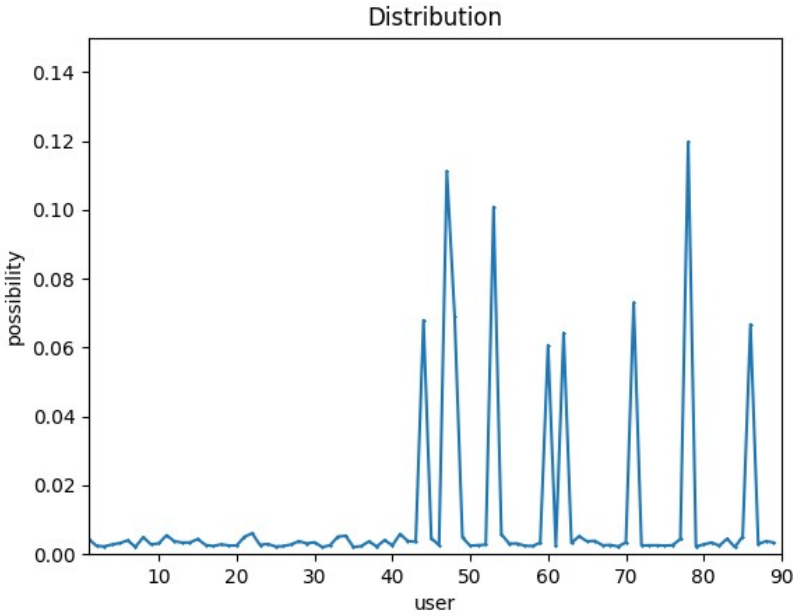
Στο τέλος της εκτέλεσης εκτυπώνεται το id του αρχικού αποστολέα και τα id των χρηστών που έγιναν **detected** στην ακόλουθη μορφή: 1, 4, 74, -1 (sender = 1, detected 4 & 74). Το -1 σημαίνει οτι δεν έγινε κανένα detection.

Αρχική γνώση

Για την δημιουργία του αρχείου users, το οποίο περιέχει τα id των sender για κάθε γύρο εκτέλεσης του πρωτοκόλλου, χρησιμοποιείται αρχικά το αρχείο corr για να

αποκλειστούν οι χρήστες που ελέγχονται από τον αντίπαλο καθώς επίσης και το distribution το οποίο περιέχει την πιθανότητα ενός χρήστη (μη corrupted) να ξεκινήσει το πρωτόκολλο.

Σε μεταγενέστερο στάδιο θα γίνει σύγκριση των αποτελεσμάτων που προκύπτουν από δύο διαφορετικές εκτελέσεις με την αρχική γνώση να είναι είτε 1) ομοιόμορφη είτε 2) κάποιοι χρήστες να έχουν μεγαλύτερη πιθανότητα από τους υπόλοιπους (εικόνα 1). Σε όλες τις περιπτώσεις που θα συγκριθούν χρησιμοποιείται γράφος με **100 χρήστες** και γίνονται **100,000 εκτελέσεις** του πρωτοκόλλου. Επίσης θα γίνει σύγκριση της ανωνυμίας που προσφέρεται σε σχέση με κάποιες παραμέτρους.



Εικόνα 1: Μη-Ομοιόμορφη Κατανομή

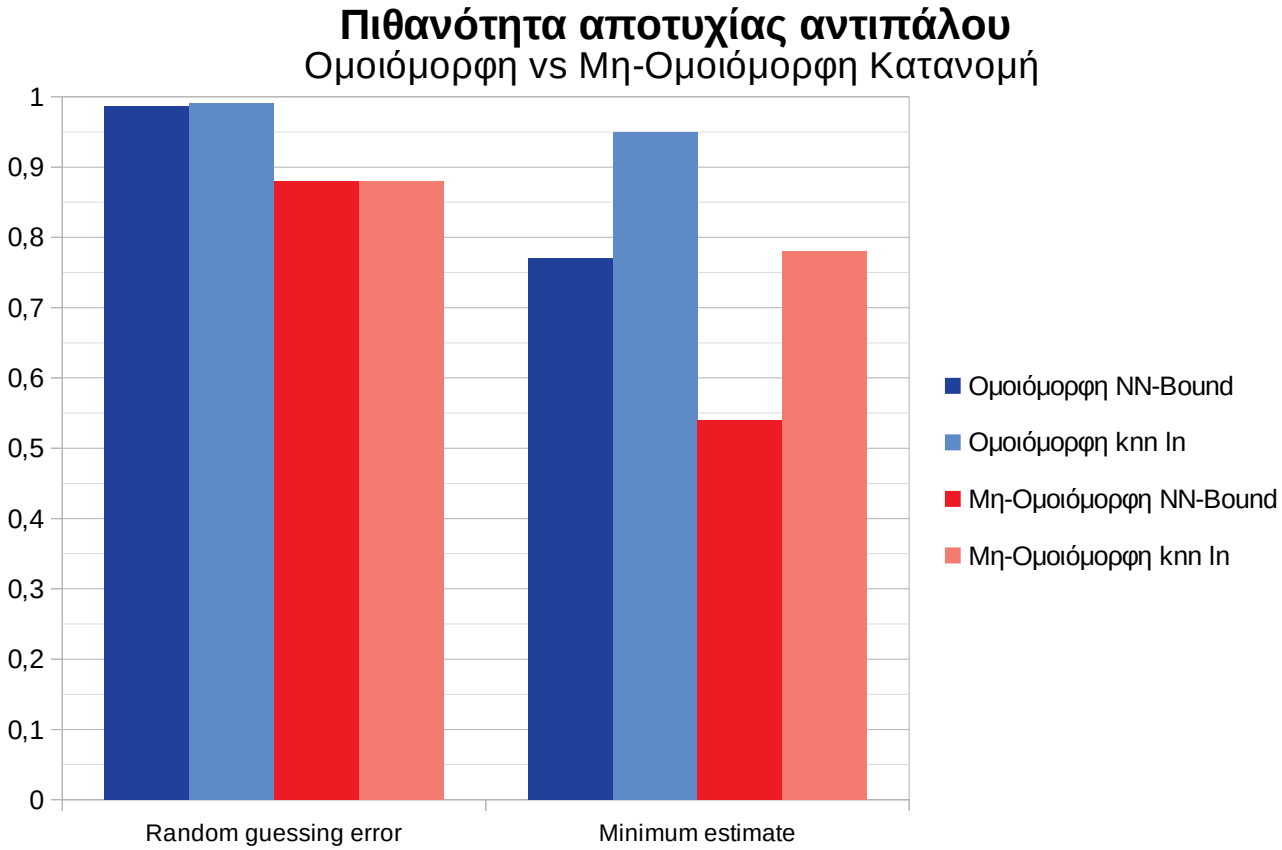
Συσχέτιση με την αρχική γνώση

Η αρχική γνώση του αντιπάλου περιέχει την κατανομή της πιθανότητας που έχει ο κάθε χρήστης ώστε να ξεκινήσει το πρωτόκολλο και να είναι ο αρχικός αποστολέας του μηνύματος.
Θα συγκρίνουμε την πιθανότητα επιτυχίας του αντιπάλου πριν και μετά την εκτέλεση του πρωτοκόλλου και για τις δύο περιπτώσεις, στους 4 διαθέσιμους αλγορίθμους που προσφέρει το F-BLEAU (Plot 1).
Αυτό που φαίνεται με μια πρώτη ματιά από την συγκεκριμένη γραφική παράσταση, είναι η μεγαλύτερη πιθανότητα αποτυχίας του αντιπάλου, είτε χωρίς την αρχική γνώση είτε με την αρχική γνώση, όταν η αρχική γνώση είναι ομοιόμορφη κατανομή πιθανότητας.
Οι υπόλοιπες παράμετροι μένουν σταθερές ως εξής:

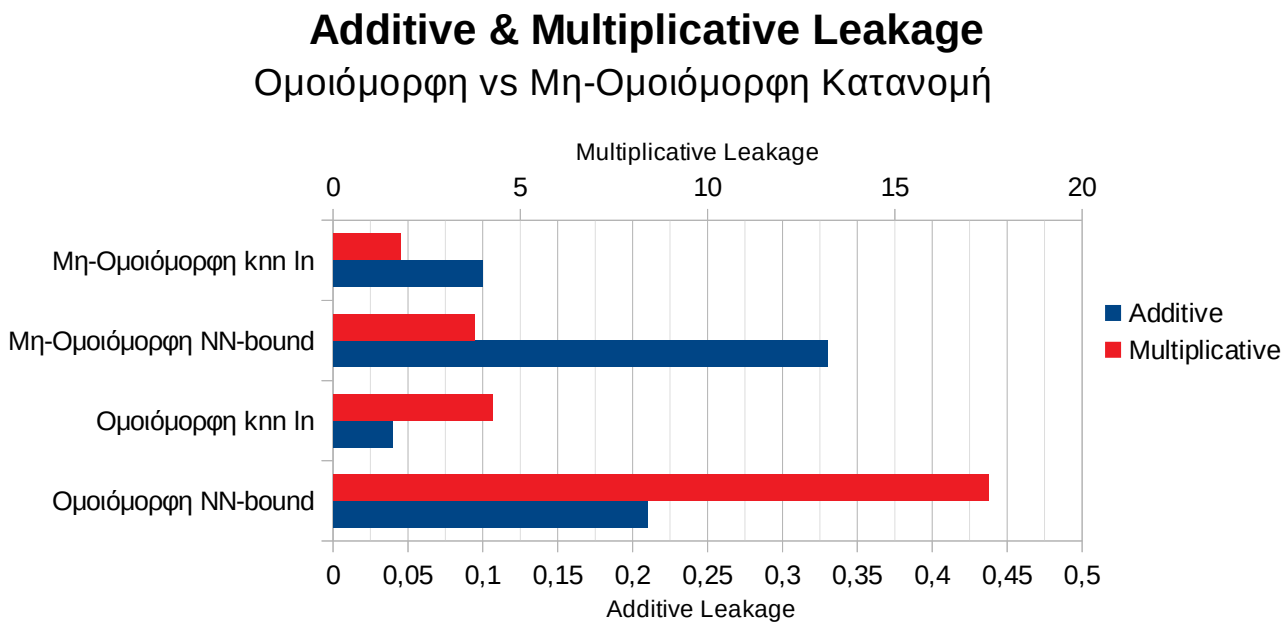
- φ = 0.75
- m = 100

- c = 10
- users = 100,000 (πόσες φορές θα εκτελεστεί το πρωτόκολλο)
- broken-paths = 2
- fix-strategy = last_honest

Επίσης, γίνεται σύγκριση της διαρροής διατηρώντας τις πιο πάνω παραμέτρους και αλλάζοντας μόνο την κατανομή πιθανότητας, δηλαδή την αρχική γνώση (Plot 2). Το συμπέρασμα που μπορεί να εξάγει με ευκολία κανείς από αυτή τη γραφική παράσταση, είναι οτι η πολλαπλασιαστική διαρροή πληροφορίας (multiplicative leakage) είναι μεγαλύτερη όταν η αρχική γνώση χαρακτηρίζεται από την ομοιόμορφη κατανομή.
Οι ίδιες μετρήσεις έγιναν και για τους 4 διαθέσιμους αλγορίθμους που προσφέρει το F-BLEAU, όμως επέλεξα να παρουσιάσω αυτούς που θα έκαναν την σύγκριση εντονότερη και πιο εμφανής.



Plot 1:Πιθανότητα αποτυχίας συναρτήσει αρχ. κατανομής



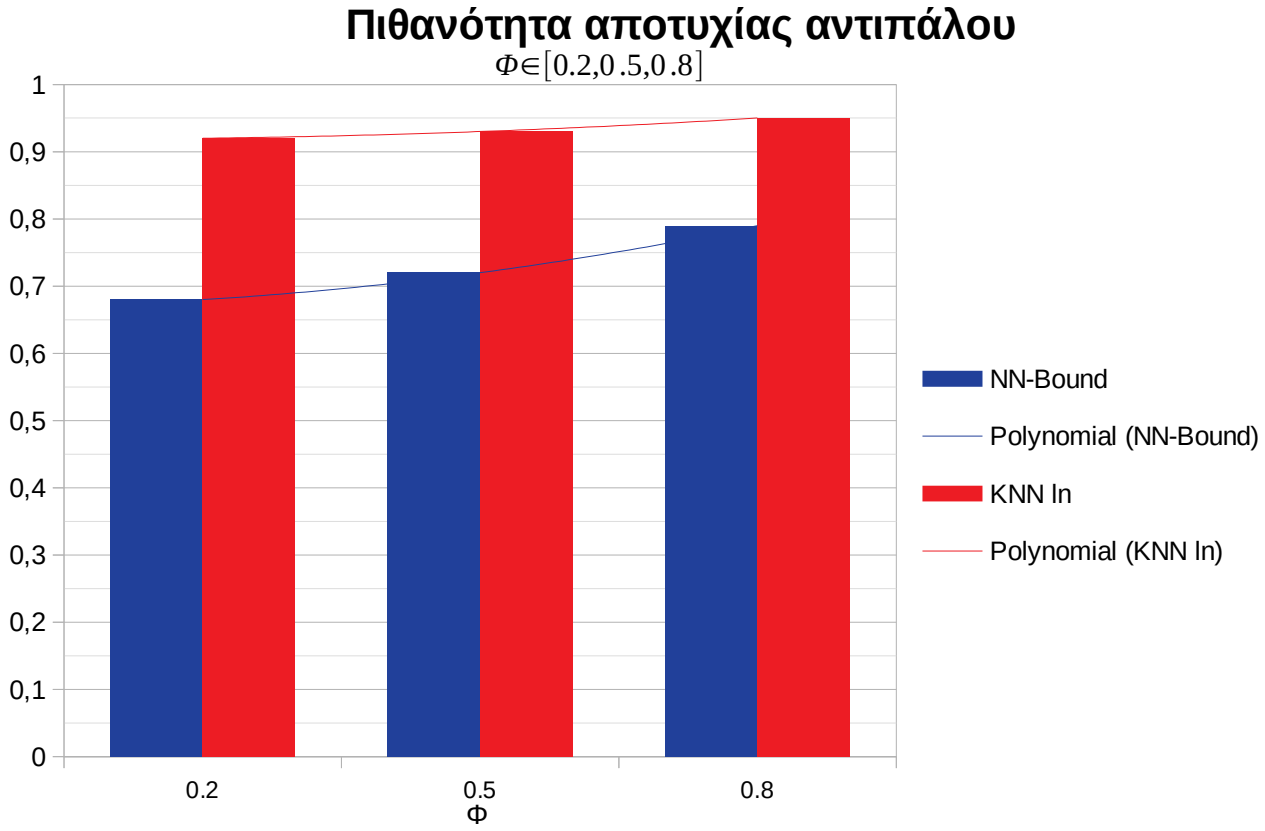
Plot 2: Διαρροή πληροφορίας συναρτήσει αρχ. κατανομής

Συσχέτιση με το Φ

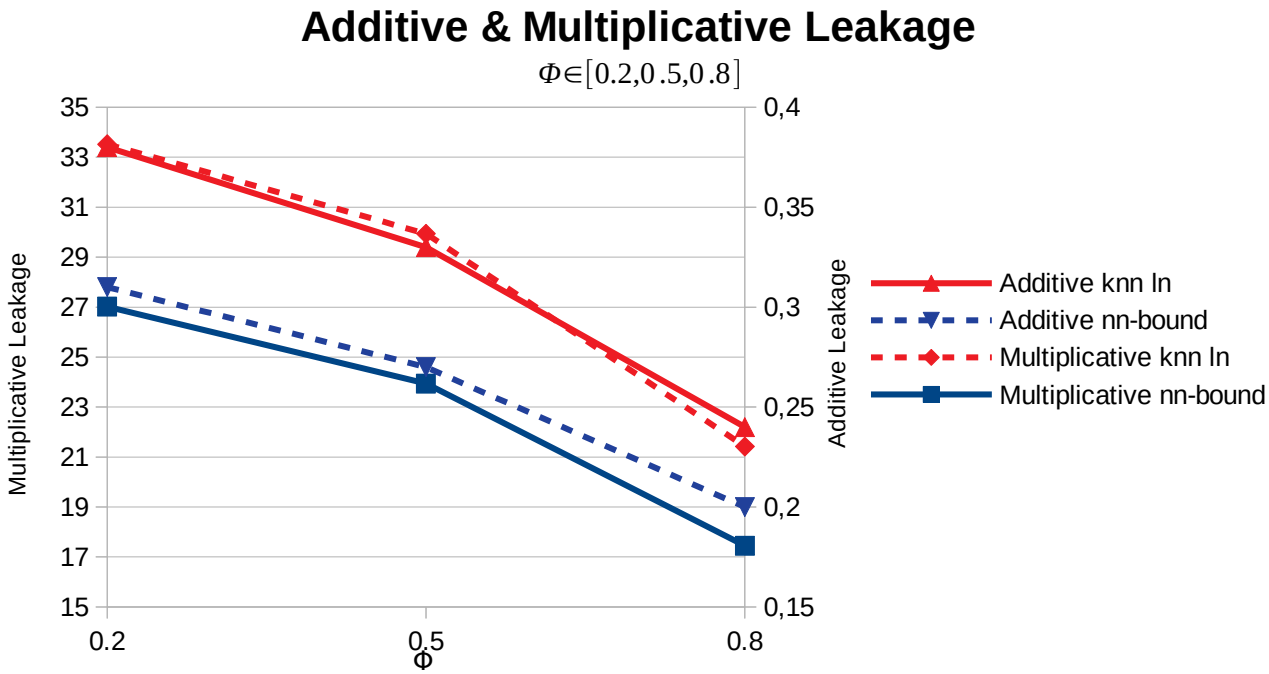
Χρησιμοποιώντας την **ομοιόμορφη κατανομή** πιθανότητας ως αρχική γνώση και τις υπόλοιπες παραμέτρους ως έχουν, θα τρέξουμε την προσομοίωση με διαφορετικές τιμές του **Φ**. Σκοπός είναι να παρατηρήσουμε πως εναλλάζεται η πιθανότητα αποτυχίας του αντιπάλου όταν έχουμε μικρό Φ (0.2), μεσαίας τάξης Φ (0.5) και όταν έχουμε μεγάλο Φ (0.8). Γίνεται επίσης σύγκριση της διαρροής πληροφορίας υπό τις ίδιες συνθήκες.

Από την αναπαράσταση της πιθανότητας αποτυχίας του αντιπάλου (Plot 3) φαίνεται ξεκάθαρα πως είναι ευθέως ανάλογη του Φ, δηλαδή όσο αυτό μεγαλώνει τόσο πιο δύσκολο γίνεται για τον αντίπαλο να μαντέψει σωστά.

Αντιθέτως, η διαρροή πληροφορίας (Plot 4) είναι αντιστρόφως ανάλογη του Φ. Αυτό σημαίνει ότι όσο μεγαλύτερο Φ έχουμε, τόσο λιγότερη διαρροή πληροφορίας συναντάμε.



Plot 3: Πιθανότητα αποτυχίας συναρτήσει Φ



Plot 4: Διαρροή πληροφορίας συναρτήσει Φ

Συσχέτιση με τον τρόπο επισκευής μονοπατιού

Διατηρώντας τις υπόλοιπες παραμέτρους σταθερές ως εξής:

- m = 100
- c = 10
- users = 100,000 (πόσες φορές θα εκτελεστεί το πρωτόκολλο)
- broken-paths = 15
- αρχική γνώση = ομοιόμορφη κατανομή πιθανότητας

παρατηρούμε στον Table 1, ότι ο τρόπος επισκευής μονοπατιών διαφοροποιεί την πιθανότητα αποτυχίας του αντιπάλου, και συνεπώς και την διαρροή πληροφορίας, μόνο σε περιπτώσεις όπου το Φ είναι χαμηλό. Στην περίπτωση που το μονοπάτι ξεκινάει από τον initiator, ο αντίπαλος έχει περισσότερες πιθανότητες να εντοπίσει τον αρχικό αποστολέα. Οι πιο κάτω μετρήσεις έγιναν με τη χρήση του αλγορίθμου nn-bound.

Φ	**Fix strategy**	Minimum Estimate	Additive Leakage	Multiplicative Leakage
0.2	Last honest	0,92	0,07	6,35
	Initiatior	0,68	0,31	27,41
0.8	Last honest	0,80	0,18	17,15
	Initiatior	0,80	0,19	17,72

Table 1: Πιθανότητα αποτυχίας & διαρροή πληροφορίας συναρτήσει fix strategy

Συσχέτιση με το πλήθος corrupted users

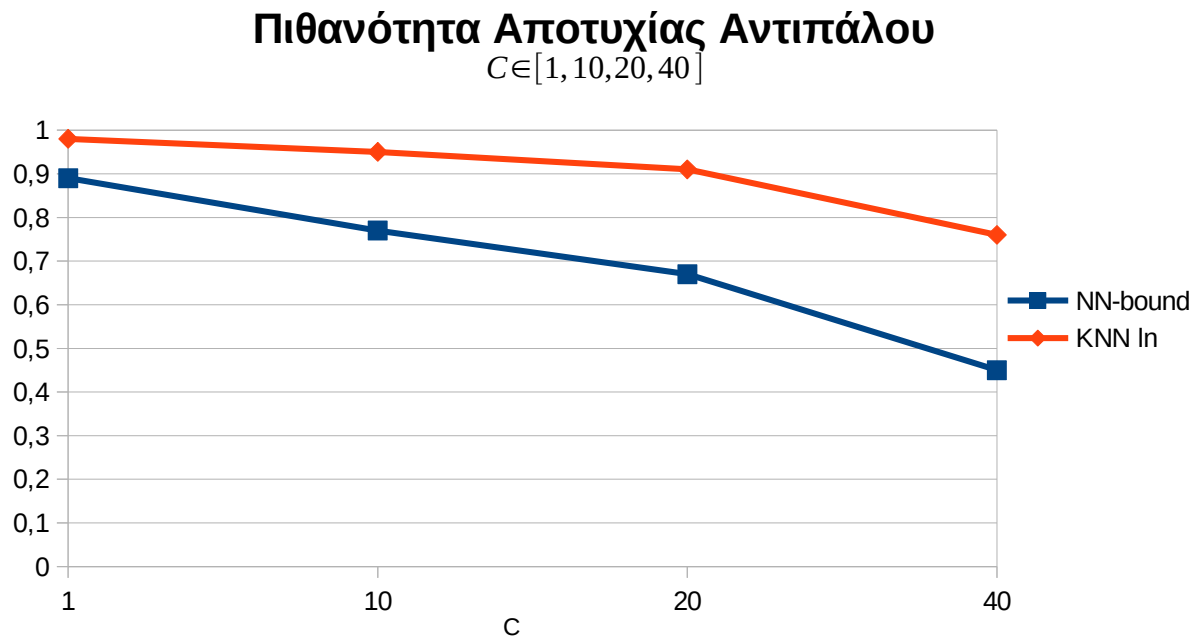
Διατηρώντας τις υπόλοιπες παραμέτρους σταθερές ως εξής:

- φ=0.75
- m = 100
- users = 100,000 (πόσες φορές θα εκτελεστεί το πρωτόκολλο)
- broken-paths = 2
- fix-strategy = last_honest
- αρχική γνώση = ομοιόμορφη κατανομή πιθανότητας

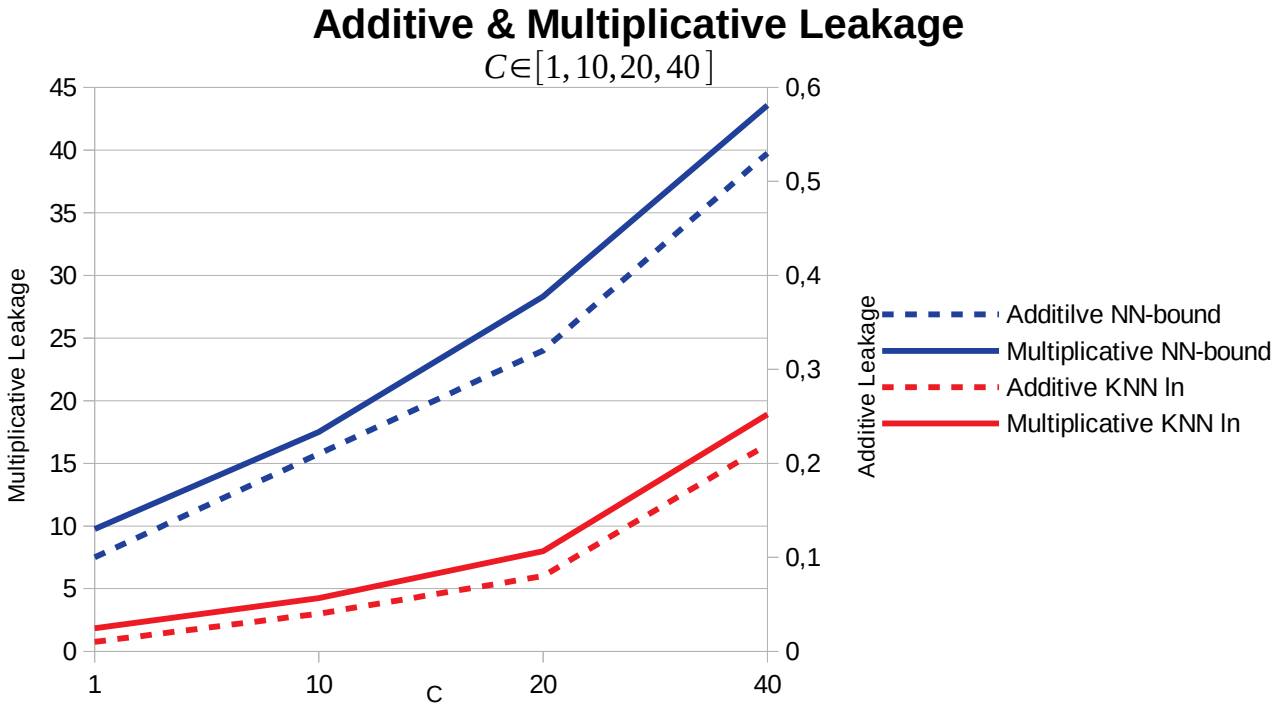
Χρησιμοποιώντας την **ομοιόμορφη κατανομή** πιθανότητας ως αρχική γνώση και τις υπόλοιπες παραμέτρους ως έχουν, θα τρέξουμε την προσομοίωση με διαφορετικές τιμές του **C** (πλήθος corrupted χρηστών). Σκοπός είναι να παρατηρήσουμε πως εναλλάζεται η πιθανότητα αποτυχίας του αντιπάλου όταν έχουμε 1, 10, 20 ή 40 corrupted χρήστες στο γράφο. Γίνεται επίσης σύγκριση της διαρροής πληροφορίας υπό τις ίδιες συνθήκες.

Από την αναπαράσταση της πιθανότητας αποτυχίας του αντιπάλου (Plot 5) φαίνεται ξεκάθαρα πως είναι αντιστρόφως ανάλογη του C, δηλαδή όσο αυτό μεγαλώνει τόσο πιο εύκολο γίνεται για τον αντίπαλο να μαντέψει σωστά.

Αντιθέτως, η διαρροή πληροφορίας (Plot 6) είναι ευθέως ανάλογη του C. Αυτό σημαίνει ότι όσο μεγαλύτερο C έχουμε, η διαρροή πληροφορίας αυξάνεται.



Plot 5: Πιθανότητα αποτυχίας συναρτήσει C



Plot 6: Διαρροή πληροφορίας συναρτήσει C

Συμπεράσματα

Τα αποτελέσματα των μετρήσεων που παρουσιάστηκαν συμφωνούν με τις σχετικές έννοιες της ποσοτικής ροής πληροφορίας. Σε όλες τις περιπτώσεις η πιθανότητα αποτυχίας του αντιπάλου βασισμένος μόνο στην αρχική γνώση είναι μεγαλύτερη από την πιθανότητα αποτυχίας του αφού δει την έξοδο. Αυτό συμφωνεί με την θεωρία που λέει οτι ο αντίπαλος δεν μπορεί να “χάσει” πληροφορία παρά μόνο να κερδίσει.

Με αυτό συμφωνούν και όλες οι μετρήσεις τόσο της Additive όσο και της Multiplicative διαρροής πληροφορίας στις οποίες συναντάμε μόνο θετικές τιμές στην πρώτη και τιμές μεγαλύτερες του 1 στην δεύτερη.

Συνοπτικά, μέσα από τις εκτελέσεις του πρωτοκόλλου και τις μετρήσεις που ακολούθησαν, προκύπτουν τα εξής συμπεράσματα:

1. Η αρχική γνώση του αντιπάλου, και συγκεκριμένα η κατανομή πιθανότητας μεταξύ των χρηστών, επηρεάζει σε μεγάλο βαθμό την πιθανότητα αποτυχίας αφού σε περίπτωση που η κατανομή δεν είναι ομοιόμορφη ο αντίπαλος έχει δυσκολότερο έργο
2. Το Φ επηρεάζει ευθέως την πιθανότητα αποτυχίας του αντιπάλου καθώς για μεγάλες τιμές ο αντίπαλος υπολογίζει τον αρχικό αποστολέα με λιγότερη επιτυχία
3. Αν επιλεγεί η επισκευή του μονοπατιού να γίνεται από τον αρχικό αποστολέα παρά από τον τελευταίο τίμιο χρήστη, τότε μόνο για χαμηλές τιμές του Φ, η πιθανότητα του αντιπάλου να υπολογίσει σωστά τον αρχικό αποστολέα μεγαλώνει. Η παρατήρηση αυτή συμφωνεί και με την θεωρία
4. Το C επηρεάζει αρνητικά την προσομοίωση, καθώς όσο αυτό αυξάνεται η πιθανότητα αποτυχίας του αντιπάλου μειώνεται και η διαρροή πληροφορίας αυξάνεται