ΠΡΟΣΤΑΣΙΑ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΣΕ ΑΣΥΡΜΑΤΑ ΔΙΚΤΥΑ ΑΝΕΚΤΙΚΑ ΣΕ ΚΑΘΥΣΤΕΡΗΣΗ

Η
ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης
του Τμήματος Πληροφορικής
Εξεταστική Επιτροπή

από τον

Βασίλειο Μπούργο

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΟ ΛΟΓΙΣΜΙΚΟ

Μάρτιος 2011

ΑΦΙΕΡΩΣΗ

Στην οικογένεια μου και στους δασκάλους μου.

ΕΥΧΑΡΙΣΤΙΕΣ

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Ευάγγελο Παπαπέτρου για την καθοριστική βοήθεια και τις συμβουλές του κατά την εκπόνηση της μεταπτυχιακής μου εργασίας. Οι γνώσεις που απέκτησα κατά τη συνεργασία μαζί του ήταν πολύτιμες.

Επιπλέον, ένα μεγάλο ευχαριστώ στον κ. Απόστολο Ζάρρα για τη φιλοξενία στο εργαστήριο του.

Επίσης, ευχαριστώ ιδιαίτερα τους γονείς μου Φώτη και Λαμπρινή για τη στήριξη που μου παρείχαν σε όλη τη διάρκεια των σπουδών μου, τόσο οικονομικά όσο και ηθικά. Χωρίς τη βοήθειά τους θα ήταν αδύνατη η περαίωσή τους.

Θερμές ευχαριστίες και σε όλους τους συναδέλφους στα εργαστήρια Middleware και NRG για τη συμπαράσταση και τις χρήσιμες υποδείξεις τους.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

ΠΕΡΙΛΗΨΗ

Βασίλειος Μπούργος του Φωτίου και της Λαμπρινής. MSc, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Μάρτιος, 2011. Προστασία ιδιωτικότητας σε ασύρματα δίκτυα ανεκτικά σε καθυστέρηση.
Επιβλέποντας: Ευάγγελος Παπαπέτρου

Η προστασία της ιδιωτικότητας είναι μια ιδιαίτερα σημαντική παράμετρος κατά τη δικτυακή επικοινωνία. Οι χρήστες ενός δικτύου έχουν την ανάγκη οι προσωπικές τους πληροφορίες να μη γίνονται διαθέσιμες σε τρίτους, χωρίς την άδειά τους. Τέτοιες πληροφορίες μπορεί να αφορούν το είδος των υπηρεσιών που χρησιμοποιούν, τους χρήστες με τους οποίους έρχονται σε επικοινωνία κ.λπ. και συνήθως εξάγονται αναλύοντας την τηλεπικοινωνιακή κίνηση ενός δικτύου.

Στην παρούσα διατριβή μελετάμε το πρόβλημα προστασίας της ιδιωτικότητας σε μια ειδική κατηγορία ασύρματων δικτύων, τα ανεκτικά σε καθυστέρηση δίκτυα (delay-tolerant networks). Τα δίκτυα αυτά αποτελούν μια απαιτητική κατηγορία δικτύων, όπου οι κόμβοι είναι κινούμενοι, ενώ δεν υπάρχει συνεχής συνδεσιμότητα μεταξύ τους με αποτέλεσμα τις μεγάλες καθυστερήσεις στην επικοινωνία. Συγκεκριμένα, εστιάζουμε στο πρόβλημα της διασφάλισης της ανωνυμίας των επικοινωνούντων κόμβων. Το πρόβλημα αυτό είναι δυσεπίλυτο για δύο λόγους. Ο πρώτος λόγος είναι ότι το ασύρματο κανάλι είναι ευάλωτο σε κακόβουλους χρήστες που «κρυφακούνε» την πληροφορία που διακινείται. Ο δεύτερος λόγος σχετίζεται με το γεγονός ότι στα δίκτυα που περιγράψαμε οι κόμβοι-χρήστες επιτελούν και τη λειτουργία της δρομολόγησης. Ωστόσο, η λειτουργία αυτή παραδοσιακά απαιτεί την αποκάλυψη της ταυτότητας του αποστολέα και του παραλήπτη και επομένως οι κακόβουλοι χρήστες μπορούν πολύ εύκολα να απειλήσουν την ανωνυμία των κόμβων που επικοινωνούν.

Η συνήθης πρακτική για την προστασία της ανωνυμίας είναι η χρήση μηχανισμών κρυπτογράφησης. Τέτοιοι μηχανισμοί είναι πολύπλοκοι και υπολογιστικά ακριβοί. Επομένως, είναι ακατάλληλοι για τις φορητές συσκευές που αποτελούν τους κόμβους των δικτύων που μελετάμε. Επιπλέον, οι τεχνικές με κρυπτογράφηση απαιτούν την εκ των προτέρων γνώση του δρομολογίου που θα ακολουθήσει ένα μήνυμα στο δίκτυο, γεγονός που είναι αδύνατο να συμβεί σε ένα δίκτυο ανεκτικό σε καθυστέρηση.

Κεντρική ιδέα στη λύση που προτείνεται είναι η χρησιμοποίηση των φίλτρων Bloom για την αναπαράσταση κρίσιμων πληροφοριών σχετικά με τη δομή του δικτύου, όπως π.χ. οι γειτονικοί κόμβοι ενός χρήστη. Οι πληροφορίες αυτές αξιοποιούνται στη συνέχεια με τη χρήση βασικών αρχών κοινωνικής δικτύωσης ώστε ένα πακέτο να δρομολογηθεί επιτυχώς στον προορισμό του χωρίς κανένας ενδιάμεσος κόμβος να είναι σε θέση να αποκαλύψει τον τελικό παραλήπτη αλλά και τον αποστολέα. Υλοποιούμε την προτεινόμενη τεχνική στον αλγόριθμο δρομολόγησης SimBet, που είναι ο πλέον γνωστός αλγόριθμος που χρησιμοποιεί τις αρχές κοινωνικής δικτύωσης. Ο νέος αλγόριθμος, που ονομάζεται SimBet-BF, προσφέρει ανώνυμη επικοινωνία με σχετικά χαμηλό κόστος, ενώ παράλληλα επιτυγχάνει δρομολόγηση εφάμιλλη (ως προς την πιθανότητα παράδοσης των μηνυμάτων αλλά και ως προς την καθυστέρηση) με τον αρχικό αλγόριθμο.

EXTENDED ABSTRACT IN ENGLISH

Bourgos Vasileios, F. MSc, Computer Science Department, University of Ioannina, Greece. March, 2011. Privacy preserving routing in opportunistic delay tolerant networks.
Thesis Supervisor: Evangelos Papapetrou

Privacy preservation is a critical issue during network communication. Network users wouldn't like their personal information to be disclosed to third parties without their permission. Such information may include the id of users that communicate with, network services they use etc. This kind of information can be extracted by analyzing network traffic.

In this thesis, we study the problem of privacy preservation in a special class of wireless networks called opportunistic networks. Opportunistic networks constitute a very challenging research area in networking. An opportunistic network consists of mobile with intermittent connectivity. This induces long delays in communication. Protecting anonymity in such an environment is hard to achieve. The wireless channel is vulnerable to malicious nodes, who can eavesdrop network information. Also, all nodes can act as routers, so they may access sensitive information during data forwarding. All known techniques for anonymity protection are using encryption mechanisms. Such mechanisms are complex and computationally expensive for the nodes participating in these networks (mobile devices).

Our approach is based on the well-known social-based algorithm, SimBet [5]. We propose the use of a special data structure, called Bloom filters. Bloom filters allow us to represent routing and other sensitive information contained in data packets in a way

that the information cannot be disclosed to intermediate nodes, while algorithm functionality is not significantly affected. Our new algorithm, called SimBet-BF, provides anonymous communication, avoiding complex cryptographic operations while simultaneously achieves routing performance very close to the original SimBet algorithm.

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1. Ασύρματη Δικτύωση - Δίκτυα ανεκτικά σε καθυστέρηση (DTN)

Τα ασύρματα δίκτυα πρωτοεμφανίστηκαν περίπου στις αρχές της δεκαετίας του 1970 και από τότε μέχρι σήμερα έχουν γίνει ιδιαίτερα δημοφιλή. Η δημιουργία τους είχε σκοπό, αρχικά, την επέκταση υπαρχόντων υποδομών ενσύρματων δικτύων. Η ανάπτυξή τους ακολουθεί γοργούς ρυθμούς, ιδιαίτερα την τελευταία δεκαετία. Νέες υπηρεσίες παρέχονται στους χρήστες, ενώ ταυτόχρονα βελτιώνονται και αναβαθμίζονται οι υποδομές. Το εύρος των εφαρμογών τους είναι πολύ μεγάλο και συνεχώς εξελισσόμενο ενώ παράλληλα, αποτελούν και έναν από τους ταχύτερα εξελισσόμενους ερευνητικούς κλάδους των τηλεπικοινωνιών και της πληροφορικής. Τα χαρακτηριστικότερα παραδείγματα εφαρμογής των ασύρματων δικτύων είναι η πρόσβαση σε υπηρεσίες διαδικτύου από χώρους όπου υπάρχει ασύρματη υποδομή (καφετέριες, αεροδρόμια, πλατείες κ.λπ.), η διασύνδεση συσκευών στο χώρο ενός σπιτιού ή μιας επιχείρησης, η πλοήγηση με χρήση GPS κ.λπ. Ωστόσο, συχνά έχουν πιο εξειδικευμένες χρήσεις όπως τα ασύρματα δίκτυα αισθητήρων ή η επικοινωνία σε καταστάσεις εκτάκτου ανάγκης (π.χ. φυσικές καταστροφές, πολεμικές επιχειρήσεις).

Τα βασικότερα πλεονεκτήματά τους είναι η δυνατότητα κίνησης των χρηστών, η απλή και γρήγορη εγκατάσταση καθώς και η εύκολη προσαρμογή και επέκτασή τους. Τα παραπάνω πλεονεκτήματα συνέβαλλαν στη ραγδαία ανάπτυξή τους. Η κινητικότητα των κόμβων εκτός από πλεονέκτημα αποτελεί ίσως το βασικότερο

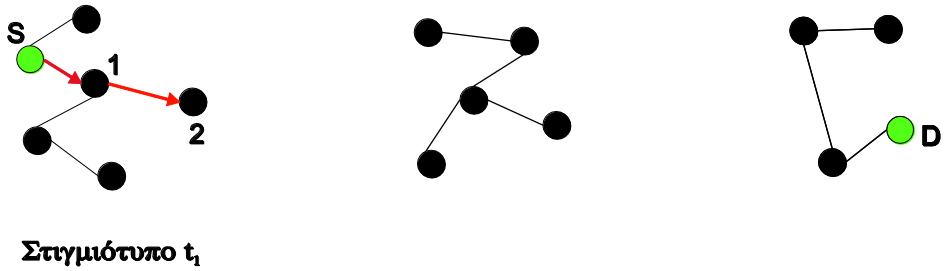
χαρακτηριστικό που διαφοροποιεί τα ασύρματα δίκτυα από τα ενσύρματα, αλλά και τις επιμέρους κατηγορίες ασύρματων δικτύων μεταξύ τους.

Τα ασύρματα δίκτυα μπορούν να κατηγοριοποιηθούν με βάση την αρχιτεκτονική τους σε δύο κατηγορίες: τα *ασύρματα δίκτυα υποδομής (infrastructure)* και τα *κατά περίπτωση δίκτυα (ad hoc)*. Στα ασύρματα δίκτυα υποδομής, το δίκτυο χωρίζεται σε περιοχές. Σε κάθε περιοχή υπάρχει ένας σταθμός βάσης *AP* (Access Point) και ένας αριθμός από ασύρματους κόμβους. Όλοι οι κόμβοι επικοινωνούν μεταξύ τους μέσω του AP. Τα AP μπορούν να συνδέονται μεταξύ τους ή/και με άλλα δίκτυα λειτουργώντας έτσι ως *πύλες*. Έτσι, οι ασύρματοι κόμβοι μπορούν να έχουν πρόσβαση στις δικτυακές υπηρεσίες μέσω των AP. Στα κατά περίπτωση δίκτυα η διάρθρωση είναι πιο απλή. Οι ασύρματοι κόμβοι είναι ισότιμοι και επικοινωνούν χωρίς τη χρήση AP. Αυτό προϋποθέτει ότι ο ένας κόμβος πρέπει να βρίσκεται στην εμβέλεια του άλλου και ότι όλοι οι κόμβοι μπορούν να λειτουργούν ως δρομολογητές. Η παραπάνω κατηγοριοποίηση σημαίνει και διαφορές στις δυνατότητες κίνησης των κόμβων. Τα AP τα οποία λειτουργούν ως δρομολογητές, δεν μπορούν να κινούνται, κάτι που δεν ισχύει στην περίπτωση της ad hoc δικτύωσης. Επίσης, οι κόμβοι πρέπει να βρίσκονται στην εμβέλεια ενός AP για να έχουν πρόσβαση σε δικτυακές υπηρεσίες. Αυτό δεν ισχύει στην ad hoc δικτύωση όπου κάθε κόμβος μπορεί να δρομολογεί δεδομένα. Γενικότερα, μπορούμε να πούμε ότι η ad hoc δικτύωση δίνει μεγαλύτερη ελευθερία κίνησης στους κόμβους.

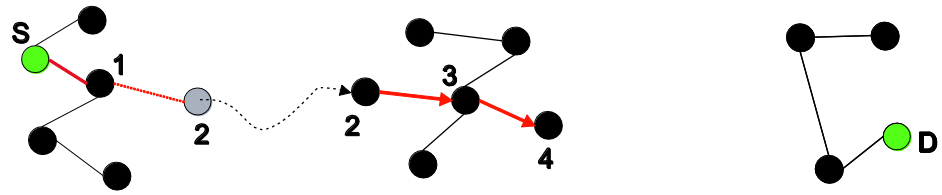
Στην κατηγορία των ad hoc ασύρματων δικτύων ιδιαίτερο ενδιαφέρον παρουσιάζει μια ειδική κατηγορία, τα *Ανεκτικά σε Καθυστέρηση Δίκτυα* (Delay Tolerant Networks - DTNs)[1]. Ένα DTN αποτελείται από ένα σύνολο κινούμενων κόμβων και πιθανόν κάποιους κόμβους που λειτουργούν ως σταθερή υποδομή. Η τοπολογία του δικτύου είναι δυναμική λόγω της κινητικότητας και υπάρχει *έλλειψη συνεχούς συνδεσιμότητας* (*intermittent connectivity*) μεταξύ των κόμβων. Κατά συνέπεια, σε κάποιο στιγμιότυπο του δικτύου, μπορεί να μην υπάρχει πλήρες μονοπάτι από έναν κόμβο-αποστολέα προς κάποιον κόμβο-παραλήπτη. Ωστόσο, τμήματα του μονοπατιού μπορεί να εμφανίζονται σε διαφορετικές χρονικές στιγμές, όπως φαίνεται στο Σχήμα

[1] Στο υπόλοιπο της διατριβής τα Ανεκτικά σε Καθυστέρηση Δίκτυα θα αναφέρονται ως DTN.

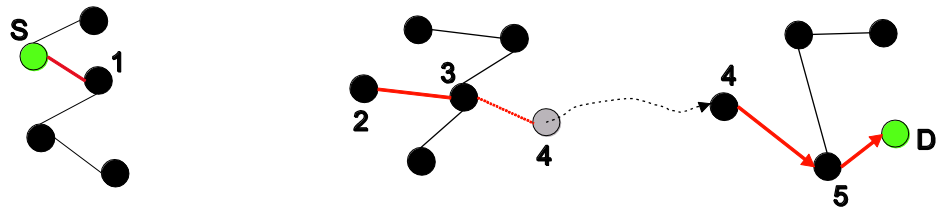
1.1. Στο σχήμα παρατηρούμε ότι ο κόμβος S θέλει να στείλει δεδομένα στον κόμβο D, όμως δεν μπορεί να το κάνει άμεσα χρησιμοποιώντας κάποιο γνωστό μονοπάτι. Συνεπώς, αναγκάζεται να προωθήσει τα δεδομένα σε ενδιάμεσους κόμβους. Έτσι, βλέπουμε ότι στο στιγμιότυπο t_1 τα δεδομένα έχουν φτάσει μέσω του κόμβου 1, στον κόμβο 2. Στη συνέχεια, ο κόμβος 2 καθώς κινείται, συναντά άλλους κόμβους οι οποίοι έχουν μεγαλύτερη πιθανότητα να παραδώσουν τα δεδομένα στον προορισμό D, οπότε τους τα προωθεί (στιγμιότυπο t_2). Τελικά, ο κόμβος 5 συναντά τον προορισμό D και του παραδίδει τα δεδομένα (στιγμιότυπο t_3).



Στιγμιότυπο t_1



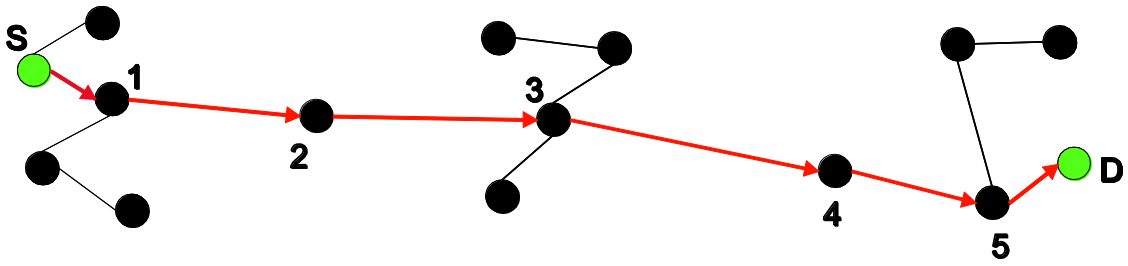
Στιγμιότυπο t_2



Στιγμιότυπο t_3

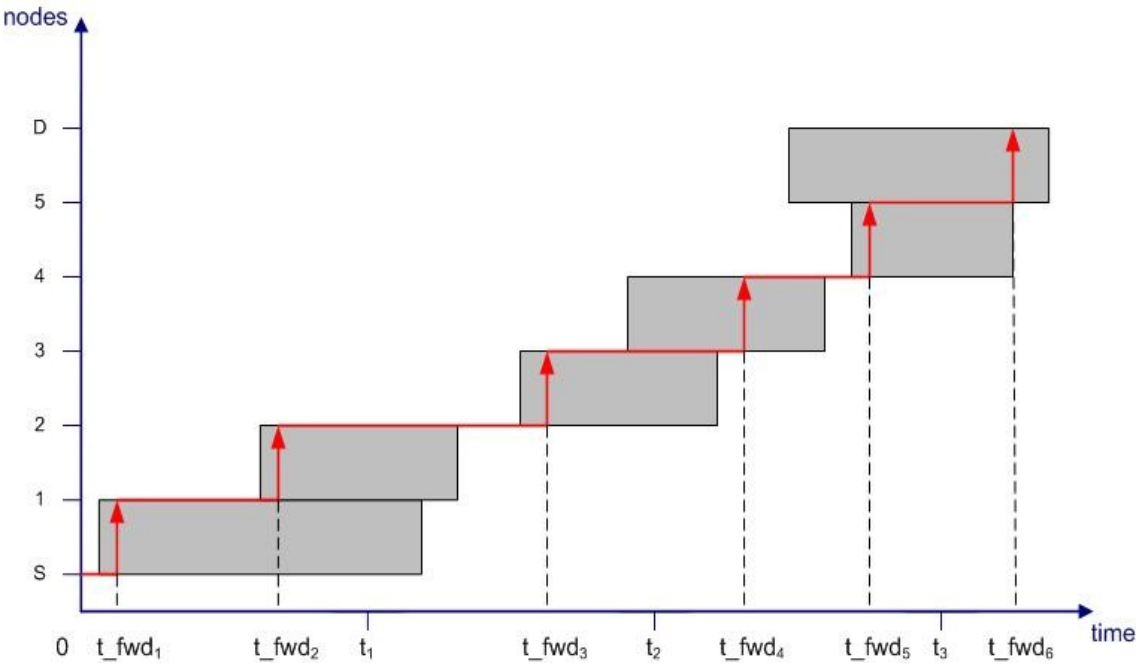
Σχήμα 1.1 Στιγμιότυπα μετάδοσης σε ένα DTN.

Η επικοινωνία, λοιπόν, σε ένα DTN γίνεται τμηματικά και το συνολικό μονοπάτι μπορούμε να το αναπαραστήσουμε συνενώνοντας τα επιμέρους μονοπάτια που σχηματίζονται στα διάφορα στιγμιότυπα (Σχήμα 1.2).



Σχήμα 1.2 Πλήρες μονοπάτι μεταξύ S,D ως συνένωση επιμέρους μονοπατιών.

Η ροή της πληροφορίας πραγματοποιείται μέσω περιστασιακών συνδέσεων μεταξύ των κόμβων, όπως φαίνεται στο Σχήμα 1.3. Όταν μια τέτοια σύνδεση είναι διαθέσιμη (π.χ. τη στιγμή t_fwd_1 ο S έχει σύνδεση με τον 1), ανάλογα με το πρωτόκολλο που χρησιμοποιείται, είναι δυνατή η προώθηση του πακέτου. Όταν δεν υπάρχει κάποια διαθέσιμη σύνδεση, ο κόμβος κρατά αποθηκευμένο το πακέτο (Σχήμα 1.3) μέχρι να εμφανιστεί κάποια σύνδεση και ανάλογα με το πρωτόκολλο που χρησιμοποιείται, να προωθηθεί το πακέτο.



Σχήμα 1.3 Απεικόνιση της προώθησης δεδομένων μέσω ευκαιριακών συνδέσεων

Στο Σχήμα 1.3 ο κόμβος 2 λαμβάνει το πακέτο από τον 1 και δεν διαθέτει κάποια ενεργή σύνδεση με κάποιον άλλο κόμβο για να προωθήσει το πακέτο. Το πακέτο, λοιπόν, μένει αποθηκευμένο στη μνήμη του. Καθώς κινείται συναντά τον κόμβο 3 και το πακέτο προωθείται στον 3 τη χρονική στιγμή t_fwd_3 και από εκεί στον 4, τη χρονική στιγμή t_fwd_4. Ο κόμβος 4, μη έχοντας κάποια ενεργή σύνδεση με άλλον κόμβο, ακολουθεί τη στρατηγική που εφάρμοσε πιο πριν ο 2, δηλ. κρατά το πακέτο μέχρι να συναντήσει κόμβους που να μπορεί να τους το προωθήσει. Τη χρονική στιγμή t_fwd_5 ο κόμβος 4 συναντά τον κόμβο 5 και του προωθεί το πακέτο και αυτός με τη σειρά του, έχοντας εντός εμβέλειας τον προορισμό D, του παραδίδει το πακέτο.

Σε τέτοια δίκτυα οι καθυστερήσεις στην επικοινωνία, λόγω της απουσίας συνδεσιμότητας, αποτελούν πολύ συχνό φαινόμενο (είναι ο λόγος για τον οποίο τους αποδόθηκε αυτή η ονομασία), καθώς οι κόμβοι μπορεί να κρατούν τα πακέτα αρκετό χρονικό διάστημα μέχρι να εμφανιστεί μια διαθέσιμη σύνδεση, ώστε να το προωθήσουν. Η από άκρο σε άκρο επικοινωνία καθίσταται ιδιαίτερα δύσκολη και πολλές φορές είναι αδύνατη, αφού ποτέ δεν είναι εγγυημένη η ύπαρξη μονοπατιού μεταξύ αποστολέα - παραλήπτη. Ένα συχνό φαινόμενο είναι να παρουσιάζονται *διαμερίσεις* (partitions), δηλ. ομάδες κόμβων να απομονώνονται από το υπόλοιπο δίκτυο. Όταν συμβαίνει αυτό, οι κόμβοι μιας διαμέρισης δεν μπορούν να επικοινωνήσουν με το υπόλοιπο δίκτυο. Για να επιλυθεί αυτό το πρόβλημα θα πρέπει το πρωτόκολλο δρομολόγησης να σχεδιαστεί κατάλληλα, ώστε να εκμεταλλεύεται όσο γίνεται καλύτερα τις ευκαιριακές συνδέσεις που παρουσιάζονται, προωθώντας τα δεδομένα ανάμεσα στα μη συνδεδεμένα τμήματα του δικτύου. Όλες οι παραπάνω ιδιαιτερότητες και προκλήσεις κάνουν τα δίκτυα αυτά τον πιο απαιτητικό τύπο δικτύων.

Η κινητικότητα των κόμβων σε ένα τέτοιο δίκτυο παίζει πολύ σημαντικό ρόλο και μπορεί να παρουσιάζει κάποια περιοδικότητα ή να είναι εντελώς στοχαστική. Στη δεύτερη περίπτωση έχουμε μια ειδική κατηγορία DTN, τα λεγόμενα *opportunistic* δίκτυα[2]. Τα opportunistic δίκτυα χαρακτηρίζονται από πλήρη έλλειψη υποδομής, ενώ

[2] Συχνά οι όροι DTN και opportunistic networks χρησιμοποιούνται για να περιγράψουν τον ίδιο τύπο δικτύου.

η κίνηση των κόμβων γίνεται εντελώς τυχαία, άρα η τοπολογία του δικτύου μεταβάλλεται στοχαστικά. Αυτό κάνει τα opportunistic δίκτυα πιο απαιτητικά και δύσκολα στη διαχείριση. Η έλλειψη υποδομής σημαίνει ότι κάθε κόμβος μπορεί να επικοινωνεί μόνο με κόμβους που βρίσκονται στην εμβέλειά του και επιπλέον, απαιτεί από κάθε κόμβο να μπορεί να ενεργεί ως δρομολογητής. Με βάση τα παραπάνω μπορούμε να πούμε ότι opportunistic δίκτυα αποτελούν την πιο γενική μορφή DTN δικτύων και την πιο απαιτητική.

Με βάση τις παραπάνω ιδιότητες βλέπουμε ότι τα opportunistic δίκτυα μοιάζουν με επέκταση των *Κινητών Κατά Περίπτωση Δικτύων* (MANETs). Η ουσιαστική διαφορά είναι η απουσία συνεχούς συνδεσιμότητας και ο τρόπος με τον οποίο αυτή αντιμετωπίζεται. Στα MANETs η λειτουργία των αλγόριθμων δρομολόγησης προϋποθέτει την ύπαρξη μονοπατιού μεταξύ αποστολέα – παραλήπτη. Αυτό σημαίνει ότι αποστολέας και παραλήπτης πρέπει να βρίσκονται στην ίδια διαμέριση. Αν αυτό δεν ισχύει, η αδυναμία εύρεσης μονοπατιού οδηγεί σε απόρριψη του πακέτου και για το λόγο αυτό οι αλγόριθμοι δρομολόγησης σε MANETs δεν μπορούν να χρησιμοποιηθούν σε DTN/opportunistic δίκτυα. Η ύπαρξη διαμερίσεων είναι πολύ συχνή στα opportunistic δίκτυα και αυτό θα έκανε την εφαρμογή αλγορίθμων για MANETs ασύμφορη για την αποδοτικότητα του δικτύου. Για να ξεπεραστεί το πρόβλημα της ύπαρξης διαμερίσεων, στα opportunistic δίκτυα ακολουθείται διαφορετική τακτική για την προώθηση των δεδομένων. Η νέα τακτική που χρησιμοποιείται είναι η *store-carry-and-forward* που είδαμε με παράδειγμα στα σχήματα 1.1 - 1.3. Αυτό σημαίνει ότι κάθε κόμβος πρέπει να διατηρεί αποθηκευμένα τα δεδομένα που έχει λάβει στο παρελθόν και να τα μεταφέρει μέχρι να παρουσιαστεί η ευκαιρία να τα προωθήσει σε κάποιον άλλο κόμβο ή να τα παραδώσει στον τελικό προορισμό. Τέτοιες ευκαιρίες παρουσιάζονται όταν συναντά άλλους κόμβους του δικτύου. Η απόφαση για προώθηση ή μη των δεδομένων εξαρτάται από το πρωτόκολλο δρομολόγησης που χρησιμοποιείται. Αναλυτική περιγραφή των γνωστότερων αλγορίθμων δρομολόγησης σε opportunistic δίκτυα θα γίνει στο κεφάλαιο 2. Από τα παραπάνω γίνεται σαφές ότι το δυναμικό περιβάλλον δημιουργεί πολλά προβλήματα επικοινωνίας, λόγω των συνεχών μεταβολών στην τοπολογία.

1.2. Αντικείμενο της Διατριβής

Η διατριβή αυτή ασχολείται με μια ενδιαφέρουσα και κρίσιμη παράμετρο της επικοινωνίας στα opportunistic δίκτυα, την *ιδιωτικότητα*. Η ιδιωτικότητα είναι μια έννοια πολύπλευρη. Γενικά, μπορούμε να πούμε ότι ο όρος ιδιωτικότητα αναφέρεται στην διαφύλαξη ευαίσθητων πληροφοριών των χρηστών ενός δικτύου από μη εξουσιοδοτημένη πρόσβαση [14]. Τέτοιες πληροφορίες μπορεί να αφορούν τη γεωγραφική θέση του κόμβου στο δίκτυο, τους προορισμούς των προς αποστολή δεδομένων του κ.λπ. Όλα τα πρωτόκολλα δρομολόγησης απαιτούν τέτοιου είδους ανταλλαγή πληροφορίας. Για παράδειγμα, σε όλα τα πρωτόκολλα πρέπει να είναι γνωστός ο προορισμός κάθε πακέτου, προκειμένου να αποφασιστεί ο επόμενος κόμβος που θα το προωθήσει και επιπλέον, πάνω σε κάθε πακέτο περιέχεται η διεύθυνση του αποστολέα. Αυτή η πληροφορία αποκαλύπτεται σε κάθε ενδιάμεσο κόμβο όταν το δίκτυο είναι τύπου ad hoc, επειδή όλοι οι κόμβοι μπορούν να είναι εν δυνάμει δρομολογητές. Συνεπώς, ένας κακόβουλος κόμβος - χρήστης που βρίσκεται πάνω στο μονοπάτι μπορεί εύκολα να αποκαλύψει τα ζεύγη επικοινωνίας των πακέτων που διακινούνται στο δίκτυο. Παρατηρούμε, λοιπόν, ότι η πληροφορία που είναι απαραίτητη για τη σωστή λειτουργία των αλγορίθμων δρομολόγησης είναι πληροφορία που παραβιάζει την ιδιωτικότητα σε ένα δίκτυο. Ουσιαστικά, λοιπόν, το πρόβλημα της εξασφάλισης ιδιωτικότητας έρχεται σε σύγκρουση με τη λειτουργία της δρομολόγησης.

Η έννοια της ιδιωτικότητας στα δίκτυα σχετίζεται συχνά με την όρο *ανωνυμία*[3]. Η προστασία της ιδιωτικότητας σε ένα δίκτυο περιλαμβάνει μηχανισμούς που αποτρέπουν την συλλογή δικτυακών πληροφοριών που σχετίζονται με ένα χρήστη. Αυτό που θέλουμε να πετύχουμε είναι η εξασφάλιση ανώνυμης επικοινωνίας, δηλ. ο κάθε κόμβος και τα δεδομένα που προέρχονται ή προορίζονται γι' αυτόν πρέπει να παραμένουν ανώνυμα. Χαρακτηριστικά παραδείγματα είναι η ανώνυμη περιήγηση στον παγκόσμιο ιστό ή η απόκρυψη πληροφοριών στις διαδικτυακές εμπορικές συναλλαγές.

---

[3] Στο υπόλοιπο της διατριβής ο όρος ιδιωτικότητα θα ταυτίζεται με την ανωνυμία

Οι στόχοι των επιθέσεων κατά της ιδιωτικότητας ποικίλουν, ανάλογα με το είδος της πληροφορίας που προσπαθεί να αποκτήσει ο κακόβουλος χρήστης. Οι περισσότερες επιθέσεις σκοπεύουν στην αποκάλυψη του ζεύγους επικοινωνίας αποστολέα-παραλήπτη. Σε αυτή την περίπτωση, ο επιτιθέμενος προσπαθεί να ανακαλύψει τον αποστολέα ή/και τον παραλήπτη ενός μηνύματος, αναλύοντας την πληροφορία που βρίσκεται πάνω στο πακέτο δεδομένων ή την τηλεπικοινωνιακή κίνηση στο δίκτυο γενικότερα. Σε άλλου τύπου επιθέσεις επιχειρείται εξαγωγή πληροφορίας σχετικά με τη δομή του δικτύου. Εδώ, συνδυάζεται πληροφορία που αφορά τις τοπικές γειτονιές των κόμβων, ώστε να αποκτηθεί ευρύτερη γνώση της τοπολογίας και των συσχετίσεων μεταξύ συγκεκριμένων κόμβων-στόχων. Παραλλαγή της παραπάνω επίθεσης αποτελεί η προσπάθεια εντοπισμού της γεωγραφικής θέσης ενός κόμβου στο δικτύου, «ακούγοντας» ποιοι κόμβοι είναι ενεργοί εντός της εμβέλειας του επιτιθέμενου. *Στην παρούσα διατριβή ασχολούμαστε με το κομμάτι της ιδιωτικότητας που αφορά την παροχή ανωνυμίας μεταξύ ενός ζεύγους επικοινωνίας (αποστολέα - παραλήπτη)*.

Υπάρχουν πολλές ερευνητικές εργασίες που σχετίζονται με την προστασία της ιδιωτικότητας τόσο σε δίκτυα ομότιμων κόμβων, όσο και σε κινητά κατά περίπτωση δίκτυα (MANETs). Ωστόσο, οι προτεινόμενες λύσεις χρησιμοποιούν κάποιο είδος κρυπτογράφησης (συμμετρική ή ασύμμετρη), κάτι που δεν είναι ιδιαίτερα αποδοτικό σε περιβάλλοντα όπως τα DTN ή τα opportunistic δίκτυα. Η κρυπτογράφηση είναι μια λειτουργία ακριβή υπολογιστικά και επιβαρύνει σημαντικά τις φορητές συσκευές που συμμετέχουν, αφού οι υπολογιστικές δυνατότητές τους είναι περιορισμένες. Τεχνικές δοκιμασμένες με επιτυχία σε δίκτυα ομότιμων κόμβων, δεν μπορούν να εφαρμοστούν σε opportunistic δίκτυα επειδή, εκτός από τις λειτουργίες κρυπτογράφησης, απαιτούν εξ' αρχής καθορισμό του μονοπατιού μεταξύ αποστολέα-παραλήπτη. Το ίδιο πρόβλημα έχουν και οι τεχνικές που σπάνε το αρχικό μήνυμα σε θραύσματα (segments) και το στέλνουν από διαφορετικά μονοπάτια. Συνεπώς, απαιτείται τελείως διαφορετική προσέγγιση για την προστασία της ιδιωτικότητας στα opportunistic δίκτυα, μιας και δεν υπάρχει καθολική γνώση για την τοπολογία. Από την άλλη πλευρά, οι εργασίες που έχουν γίνει σε opportunistic δίκτυα είναι λιγοστές και στηρίζονται και αυτές σε κρυπτογράφηση ή/και εισάγουν υποδομή η οποία μπορεί να θεωρηθεί αξιόπιστη όσον αφορά την διαχείριση ευαίσθητης πληροφορίας.

Στόχος της παρούσας διατριβής είναι να προτείνουμε έναν πιο αποδοτικό τρόπο για την προστασία της ιδιωτικότητας σε opportunistic δίκτυα. Η λύση μας εστιάζει στους αλγορίθμους που λειτουργούν με αρχές κοινωνικής δικτύωσης. Η κοινωνική δικτύωση αποτελεί μια διαφορετική προσέγγιση σχετικά με τον τρόπο λειτουργίας των opportunistic δικτύων. Σύμφωνα με αυτή, οι κοινωνικές σχέσεις μεταξύ των ανθρώπων στην καθημερινότητά τους μεταφέρονται και στη συμπεριφορά τους ως κόμβων ενός opportunistic δικτύου. Αυτή η προσέγγιση αντανακλά σε μεγάλο βαθμό την πραγματικότητα, αφού η χρήση ασύρματων φορητών συσκευών φέρνει σε επαφή ανθρώπους που συνδέονται με διάφορες κοινωνικές σχέσεις (φίλοι, συνάδελφοι κ.λπ.). Αυτό καθορίζει ως ένα βαθμό τον τρόπο που κινούνται και συμπεριφέρονται οι κόμβοι-χρήστες σε ένα τέτοιο δίκτυο. Η κοινωνική πτυχή των opportunistic δικτύων έχει αποδειχθεί πειραματικά με χρήση φορητών συσκευών, οι οποίες κατέγραφαν τις επαφές των χρηστών που συμμετείχαν στο πείραμα. Παρατηρήθηκε, λοιπόν, ότι η κίνηση και η επικοινωνία μεταξύ των κόμβων δεν ήταν εντελώς στοχαστική, άλλα σχετιζόταν άμεσα με τις κοινωνικές σχέσεις που είχαν οι χρήστες στην καθημερινή τους ζωή.

Με βάση την παραπάνω προσέγγιση, στηριχθήκαμε σε έναν από τους πιο γνωστούς αλγορίθμους κοινωνικής δικτύωσης, τον SimBet [5]. Στον αλγόριθμο αυτό, η πληροφορία που διακινείται στα πακέτα περιλαμβάνει, εκτός από τον αποστολέα και τον παραλήπτη του μηνύματος (κάτι που ισχύει σε όλους τους αλγορίθμους), πληροφορίες σχετικές με την τοπική γειτονιά κάθε κόμβου. Αυτήν την πληροφορία θέλουμε να προστατεύσουμε από τους κακόβουλους χρήστες. Η χρήση κάποιου κρυπτογραφικού μηχανισμού δεν ενδείκνυται, αφού σε αυτή την περίπτωση θα έπρεπε σε κάθε ενδιάμεσο κόμβο να πραγματοποιείται αποκρυπτογράφηση και εκ νέου κρυπτογράφηση της πληροφορίας. Αντί αυτού, προτείνουμε τη χρήση μιας ειδικής δομής δεδομένων, των *φίλτρων Bloom* [1]. Τα φίλτρα Bloom μας επιτρέπουν την αναπαράσταση ενός συνόλου με χρήση ενός δυαδικού μονοδιάστατου πίνακα. Αυτό είναι ιδιαίτερα χρήσιμο, αφού μας επιτρέπει να αναπαραστήσουμε την πληροφορία που ανταλλάσσουν οι κόμβοι κατά την δρομολόγηση με τέτοιο τρόπο, ώστε να προστατεύεται από κακόβουλους χρήστες και ταυτόχρονα να μπορεί να χρησιμοποιηθεί σχεδόν αυτούσια για τη σωστή λειτουργία του αλγορίθμου. Έτσι,

εξασφαλίζουμε την ιδιωτικότητα, διατηρώντας παράλληλα την επίδοση του αλγορίθμου σε πολύ καλά επίπεδα.

Ο νέος αλγόριθμος, που ονομάζεται SimBet-BF, αποφεύγοντας τη χρήση κρυπτογράφησης, αποτελεί ιδανική λύση για φορητές συσκευές με περιορισμένες δυνατότητες. Πειραματικά, αποδεικνύεται ότι η απόδοσή του τόσο σε επίπεδο επιτυχούς παράδοση πακέτων, όσο και στην καθυστέρηση παράδοσης, είναι εφάμιλλη με τον αρχικό αλγόριθμο, με ελάχιστες διαφοροποιήσεις.

1.3. Δομή της Διατριβής

Η διατριβή έχει την παρακάτω διάρθρωση. Στο Κεφάλαιο 2 γίνεται μια κατηγοριοποίηση των σημαντικότερων αλγόριθμοι δρομολόγησης σε DTN/opportunistic και μια περιγραφή της λειτουργίας τους, με έμφαση στους αλγορίθμους κοινωνικής δικτύωσης, που αποτελούν τη βάση της διατριβής. Έπειτα, αναλύεται η έννοια της ιδιωτικότητας σε δίκτυα, η σημασία της και οι δυσκολίες εξασφάλισής της σε περιβάλλοντα κινούμενων κόμβων. Επίσης, παρουσιάζονται οι σχετικές εργασίες πάνω στο αντικείμενο μελέτης. Στη συνέχεια, στο Κεφάλαιο 3, περιγράφεται αναλυτικά η λύση που προτείνεται, η οποία βασίζεται στον αλγόριθμο SimBet. Παράλληλα, μελετάμε τι αντίκτυπο έχει η προστασία της ιδιωτικότητας στην αποδοτικότητα της δρομολόγησης. Στο Κεφάλαιο 4, αρχικά παρουσιάζεται το περιβάλλον προσομοίωσης και οι μετρικές που χρησιμοποιήθηκαν για την αξιολόγηση της λύσης. Έπειτα, παραθέτονται τα πειραματικά αποτελέσματα και ακολουθεί ο σχολιασμός τους. Τέλος, στο Κεφάλαιο 5 συνοψίζονται τα συμπεράσματα που προκύπτουν από την παρούσα διατριβή και πιθανές μελλοντικές επεκτάσεις.

ΚΕΦΑΛΑΙΟ 2. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ - ΣΧΕΤΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

Στο κεφάλαιο αυτό παρουσιάζονται οι βασικές αρχές και αλγόριθμοι δρομολόγησης σε opportunistic δίκτυα, καθώς και μια εισαγωγή στην έννοια της ιδιωτικότητας, έτσι όπως αυτά καταγράφονται στη βιβλιογραφία. Αρχικά, παρουσιάζουμε μια κατηγοριοποίηση των αλγορίθμων, περιγράφοντας συνοπτικά τον τρόπο λειτουργίας τους, εστιάζοντας στους αλγορίθμους κοινωνικής δικτύωσης. Έπειτα, αναλύουμε την έννοια της ιδιωτικότητας και τις πτυχές της και παρουσιάζουμε τις λύσεις που έχουν προταθεί μέχρι στιγμής για την προστασία της.

2.1. Δρομολόγηση σε DTN

Η δρομολόγηση σε DTN δίκτυα παρουσιάζει, όπως είδαμε, πολλές προκλήσεις και δυσκολίες, λόγω της φύσης αυτών των δικτύων. Παρακάτω παρουσιάζουμε τις βασικότερες κατηγορίες αλγορίθμων δρομολόγησης σε DTN δίκτυα και ένα χαρακτηριστικό παράδειγμα για καθεμιά.

2.1.1. Κατηγορίες αλγόριθμων δρομολόγησης

Οι αλγόριθμοι δρομολόγησης σε opportunistic δίκτυα μπορούν γενικά να χωριστούν στις τρεις παρακάτω κατηγορίες: αλγόριθμοι βασισμένοι στην πλημμύρα (flooding-based), αλγόριθμοι βασισμένοι σε ιστορικό (history-based) και αλγόριθμοι κοινωνικής δικτύωσης (social-based).

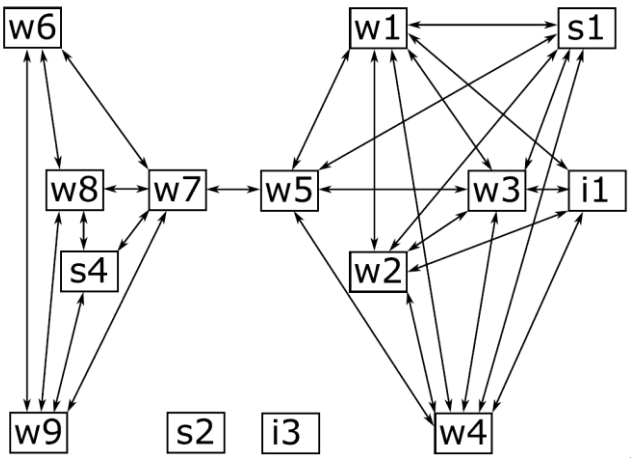
Οι αλγόριθμοι πλημμύρας αποτελούν την πιο απλή προσέγγιση για τη δρομολόγηση σε opportunistic δίκτυα. Ο πιο χαρακτηριστικός αλγόριθμος της κατηγορίας είναι ο Epidemic [19]. Σύμφωνα με αυτόν, κάθε κόμβος παράγει αντίγραφα των προς αποστολή μηνυμάτων και τα προωθεί σε όσους κόμβους συναντά στο δίκτυο. Αν κάποιος κόμβος έχει ήδη αντίγραφο του μηνύματος, τότε δε γίνεται προώθηση. Ο αλγόριθμος Epidemic επιβαρύνεται το δίκτυο με πολλά αντίγραφα των αρχικών μηνυμάτων και αυτό επιβαρύνει τόσο τη μνήμη των κόμβων, όσο και το εύρος ζώνης του δικτύου. Υπάρχουν, ωστόσο, πολλές παραλλαγές του Epidemic όπως ο PREP [16], ο Spray-and-wait [18] κ.λπ. που προσπαθούν να περιορίσουν το πλήθος των αντιγράφων, θέτοντας όριο στον αριθμό των αλμάτων (hops) ή/και στο χρόνο ζωής (ttl) ενός μηνύματος.

Οι αλγόριθμοι που βασίζονται σε ιστορικό επαφών εστιάζουν στην καλύτερη διαχείριση των πόρων. Η ιδέα βασίζεται στη χρήση του ιστορικού συναντήσεων με τους κόμβους του δικτύου. Η κεντρική ιδέα είναι ότι αν ένας κόμβος έχει συναντήσει τον κόμβο-προορισμό ενός πακέτου πολλές φορές στο παρελθόν, είναι πολύ πιθανόν να τον συναντήσει και στο μέλλον. Ο χαρακτηριστικότερος αλγόριθμος αυτής της κατηγορίας είναι ο PROPHET [11]. Ο αλγόριθμος αυτός χρησιμοποιεί σε κάθε κόμβο το ιστορικό των επαφών για να εξάγει μια πιθανότητα επιτυχούς παράδοσης για κάθε γνωστό προορισμό. Όταν δυο κόμβοι συναντιούνται η παραπάνω πληροφορία ανταλλάσσεται. Όποιος από τους δυο έχει τη μεγαλύτερη πιθανότητα επιτυχούς παράδοσης αναλαμβάνει να προωθήσει το πακέτο.

Οι αλγόριθμοι κοινωνικής δικτύωσης στηρίζονται στην ιδέα ότι οι κοινωνικές σχέσεις που ισχύουν στην ανθρώπινη πραγματικότητα αντικατοπτρίζονται στη συμπεριφορά των χρηστών στα πλαίσια της λειτουργίας ενός δικτύου. Οι αποφάσεις δρομολόγησης λαμβάνονται με βάση μετρικές που προέρχονται από το χώρο της Κοινωνικής Ανάλυσης και χαρακτηρίζουν το πόσο σημαντική θέση κατέχει ένας κόμβος στο δίκτυο. Οι πιο γνωστοί αλγόριθμοι κοινωνικής δικτύωσης είναι ο Bubble Rap [8] και ο SimBet [5]. Η παρούσα διατριβή εστιάζει στους αλγορίθμους κοινωνικής δικτύωσης (στηρίζεται στον SimBet) με σκοπό να παρέχει έναν αποτελεσματικό τρόπο για την προστασία της ιδιωτικότητας σε περιβάλλοντα κοινωνικής δικτύωσης.

2.1.2. Κοινωνική δικτύωση σε opportunistic δίκτυα

Η κοινωνική δικτύωση αποτελεί μια ενδιαφέρουσα προσέγγιση όσον αφορά τα opportunistic δίκτυα. Η ιδέα στην οποία βασίζεται είναι ότι οι κοινωνικές σχέσεις που αναπτύσσουν οι χρηστές στην καθημερινότητά τους (συνάδελφοι, φίλοι κ.λπ.) εξακολουθούν να ισχύουν όταν αυτοί αποτελούν κόμβους ενός δικτύου. Είναι σαφές ότι χρήστες που σχετίζονται π.χ. με μια επαγγελματική σχέση είναι λογικό να έχουν συχνές επαφές στην καθημερινότητά τους. Οι κοινωνικές επαφές τους συνεπάγονται και επαφές μεταξύ των συσκευών τους όταν αποτελούν χρήστες ενός δικτύου. Οι σχέσεις αυτές μπορούν να εξαχθούν παρατηρώντας τις αλληλεπιδράσεις τους κατά τη δικτυακή επικοινωνία. Κόμβοι ενός δικτύου που συναντιούνται συχνά σημαίνει ότι πιθανότατα έχουν μια κοινωνική σχέση και στην πραγματικότητα. Η συσχέτιση δυο κόμβων ονομάζεται *επαφή* (contact). Για να έχουμε επαφή μεταξύ δύο κόμβων απαραίτητη προϋπόθεση είναι οι κόμβοι αυτοί να έχουν υπάρξει γείτονες στο παρελθόν. Το κριτήριο αυτό δεν είναι σαφώς ορισμένο, αφού η ύπαρξη ενός κόμβου στη γειτονιά ενός άλλου δεν συνεπάγεται ότι υπάρχει πάντα κάποια κοινωνική σχέση μεταξύ τους (π.χ. μπορεί να έχουν συναντηθεί τυχαία). Ο καθορισμός ενός σωστού κριτηρίου για το πότε έχουμε επαφή μεταξύ δύο κόμβων είναι ένα ερευνητικό πρόβλημα με το οποίο δεν θα ασχοληθούμε στην παρούσα διατριβή. Αυτές οι σχέσεις, λοιπόν, μπορούν να αναπαρασταθούν με ένα γράφημα που ονομάζεται *κοινωνικός γράφος* (social graph), όπου οι κόμβοι είναι οι χρήστες και οι ακμές οι σχέσεις που αναπτύσσουν. Ο κοινωνικός γράφος, λοιπόν, αποτελεί μια αναπαράσταση όλων των επαφών μεταξύ των κόμβων του δικτύου.



Σχήμα 2.1 Παράδειγμα κοινωνικού δικτύου

Η δρομολόγηση που στηρίζεται στην ιδέα της κοινωνικής δικτύωσης χρησιμοποιεί μετρικές οι οποίες εξάγονται από τις συνδέσεις του κοινωνικού γράφου. Οι πιο γνωστές μετρικές είναι το *centrality* ενός κόμβου v και το *similarity* ενός κόμβου v με έναν άλλο κόμβο u [5]. Το centrality εκφράζει το πόσο σημαντικός είναι ένας κόμβος για το δίκτυο. Κόμβοι με μεγάλες τιμές centrality συνεισφέρουν περισσότερο στην συνεκτικότητα του δικτύου από άλλους με μικρότερες τιμές. Έτσι, ένας κόμβος με μεγάλο centrality, όπως ο w5 στο Σχήμα 2.1, έχει περισσότερες πιθανότητες να παραδώσει τα δεδομένα στον προορισμό. Η μετρική του centrality ενός κόμβου έχει πολλές παραλλαγές. Εδώ παρουσιάζουμε τη μετρική *betweenness centrality*, η οποία χρησιμοποιείται στην παρούσα διατριβή. Το betweenness centrality για έναν κόμβο v εκφράζει σε πόσα συντομότερα μονοπάτια μεταξύ ενός ζεύγους κόμβων u,w συμμετέχει ο v ως ενδιάμεσος κόμβος. Αν $g_{j,k}$ είναι το πλήθος των συντομότερων μονοπατιών μεταξύ οποιουδήποτε ζεύγους κόμβων j, k του δικτύου και $g_{j,k}(p)$ είναι το πλήθος των συντομότερων μονοπατιών μεταξύ των j, k στα οποία περιλαμβάνεται ο κόμβος p, τότε το betweenness centrality του κόμβου p ορίζεται από την εξίσωση 2.1

$$C_B(p) = \sum_{j=1}^{N} \sum_{k=1}^{j-1} \frac{g_{jk}(p)}{g_{jk}} \qquad \text{Εξ. 2.1}$$

Άρα, στο Σχήμα 2.1, το betweenness centrality του κόμβου w8 είναι 1/3. Το similarity ενός κόμβου v με έναν κόμβο u εκφράζει το πλήθος των κοινών επαφών τους. Αν ο κόμβος v έχει πολλές κοινές επαφές με τον u, τότε είναι πολύ πιθανόν να τον συναντήσει στο μέλλον και να του παραδώσει δεδομένα που προορίζονται γι' αυτόν. Επίσης, η ύπαρξη πολλών κοινών επαφών μεταξύ ενός κόμβου v με έναν κόμβο - προορισμό σημαίνει ότι η πιθανότητα να συναντήσει ο v κάποιον από αυτούς είναι μεγάλη. Συνεπώς, οι πολλές κοινές επαφές ενός κόμβου με τον προορισμό σημαίνει πρακτικά ότι το πακέτο έχει περισσότερες πιθανότητες να φτάσει ένα άλμα πιο κοντά στον προορισμό. Στο Σχήμα 2.1 παρατηρούμε ότι το similarity του w8 με τον w5 είναι 1, αφού ο w8 έχει έναν κοινό γείτονα με τον w5, τον κόμβο w7. Οι αλγόριθμοι που βασίζονται στην κοινωνική δικτύωση φαίνεται να έχουν πολύ καλή απόδοση στην πράξη, εξασφαλίζοντας μεγάλες πιθανότητες επιτυχούς παράδοσης των πακέτων. Παράλληλα, η επιβάρυνση για το δίκτυο και τους κόμβους κυμαίνεται σε

πολύ χαμηλότερα επίπεδα σε σχέση με άλλα πρωτόκολλα. Παρακάτω περιγράφουμε τον αλγόριθμο SimBet, στον οποίο βασίζεται η παρούσα διατριβή.

2.1.3. Ο αλγόριθμος SimBet

Ο αλγόριθμος SimBet χρησιμοποιεί τις μετρικές betweenness και similarity για να αποφασίσει ποιος κόμβος θα αναλάβει την προώθηση ενός πακέτου. Κάθε κόμβος κρατά τις επαφές του σε μια λίστα (encounter list). Η λίστα αυτή βοηθά στον υπολογισμό των παραπάνω μετρικών, καθώς οι μετρικές αυτές εξάγονται με βάση τις επαφές κάθε κόμβου και των γειτόνων του. Όταν δύο κόμβοι συναντιούνται, προκειμένου να αποφασίσουν ποιος κόμβος θα προωθήσει ποια πακέτα, συγκρίνουν τις τιμές των μετρικών τους. Για την ακρίβεια, η σύγκριση γίνεται αφού οι δυο μετρικές συνδυαστούν γραμμικά (δίνοντας τη μετρική SimBetUtil [5]). Ο συνδυασμός των μετρικών γίνεται ως εξής: έστω Α και Β οι δύο κόμβοι που επικοινωνούν με $Sim_A(d)$, $Sim_B(d)$ οι τιμές του similarity που διαθέτουν για έναν προορισμό d και Bet_A, Bet_B οι τιμές του betweenness. Ο Α θα υπολογίσει το SimBetUtil με βάση τους παρακάτω τύπους:

$$SimUtil_A(d) = \frac{Sim_A(d)}{Sim_A(d) + Sim_B(d)} \qquad \text{Εξ. 2.2}$$

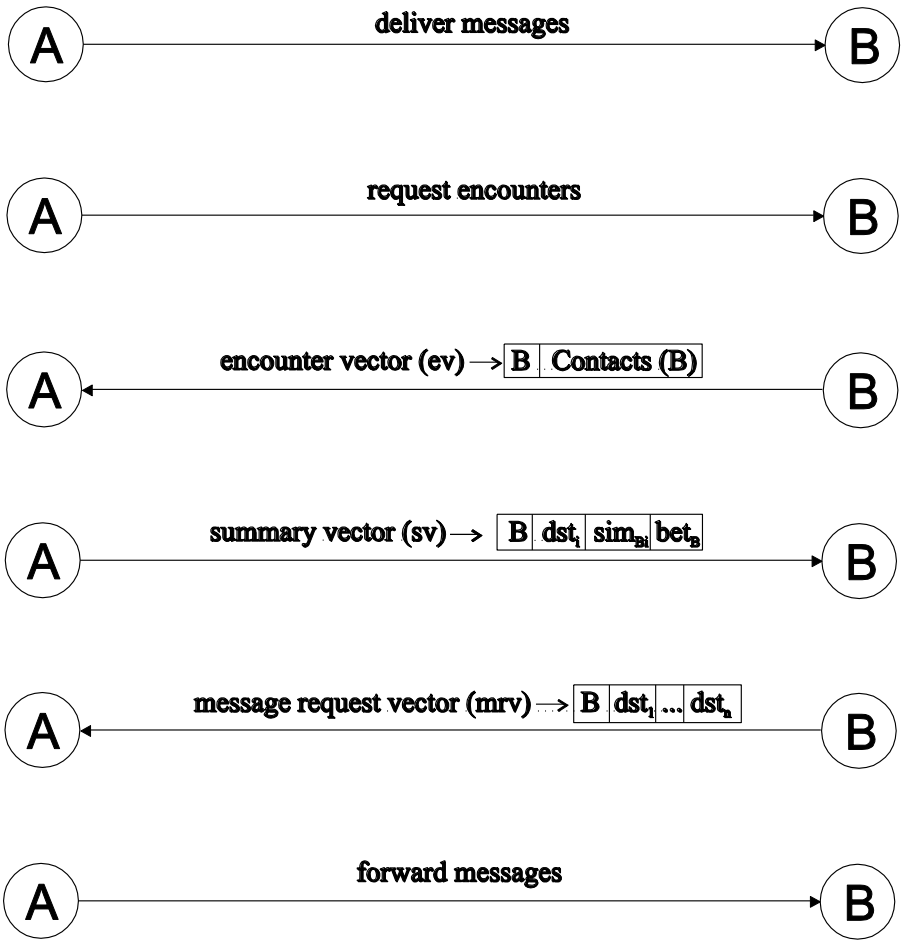
$$BetUtil_A = \frac{Bet_A}{Bet_A + Bet_B} \qquad \text{Εξ. 2.3}$$

$$SimBetUtil_A(d) = a \cdot SimUtil_A(d) + b \cdot BetUtil_A \qquad \text{Εξ. 2.4}$$

όπου α και b τίθενται ίσα με 0.5 δηλ. οι δύο μετρικές λαμβάνονται το ίδιο υπ' όψιν κατά τον υπολογισμό του SimBetUtil. Όποιος από τους κόμβους έχει μεγαλύτερη τιμή SimBetUtil αναλαμβάνει την περαιτέρω προώθηση του πακέτου.

Πιο συγκεκριμένα το πρωτόκολλο έχει ως εξής: όταν ένας κόμβος Α συναντά ένα νέο γείτονα Β, αρχικά, του παραδίδει όσα από τα αποθηκευμένα μηνύματά του έχουν τον

Β ως προορισμό. Η αναγνώριση του νέου γείτονα πραγματοποιείται με ανταλλαγή μηνυμάτων «hello». Κάθε κόμβος περιοδικά εκπέμπει ένα μήνυμα «hello» με σκοπό να γίνει αντιληπτός από τους κόμβος εντός της εμβέλειάς του. Στη συνέχεια, αφού ο Α παραδώσει όσα μηνύματα προορίζονται για τον Β, του στέλνει ένα μήνυμα με το οποίο ζητά από τον Β να του στείλει τη λίστα των κόμβων που έχει συναντήσει μέχρι στιγμής (request encounters message).



Σχήμα 2.2 Ανταλλαγή πληροφορίας στον αλγόριθμο SimBet

Η λίστα αυτή περιλαμβάνει όλες τις επαφές του Β. Όταν λάβει την παραπάνω λίστα (encounter vector), με βάση το περιεχόμενό της (επαφές του Β), υπολογίζει το betweenness και το similarity με τους κόμβους-προορισμούς των πακέτων που διαθέτει. Ουσιαστικά, ο Α ελέγχει αν υπάρχουν επαφές του που να περιέχονται στη λίστα επαφών του Β. Αν μια επαφή του Α (π.χ. ο κόμβος C) δεν υπάρχει στις επαφές του Β, αυτό αυξάνει το betweenness του Α, αφού είναι ενδιάμεσος κόμβος σε

συντομότερο μονοπάτι (2 αλμάτων) μεταξύ του B και του C. Επίσης, από την ίδια λίστα επαφών μπορεί να υπολογίσει και πιθανόν να ανανεώσει το πλήθος των κοινών επαφών του με κάθε προορισμό, δηλ. το similarity του. Αφού ολοκληρωθεί και ο υπολογισμός των μετρικών, ο A στέλνει στον B μια λίστα (summary vector message) που περιέχει όλους τους προορισμούς των πακέτων που διαθέτει, μαζί με τις τιμές betweenness και similarity που υπολόγισε. Όταν ο B λάβει την λίστα summary vector υπολογίζει για κάθε προορισμό που περιέχεται στη λίστα τις δικές του τιμές betweenness και similarity και εξάγει τη μετρική SimBetUtil. Υστέρα, συγκρίνει τις τιμές που υπολόγισε με αυτές που του έστειλε ο A και τους προορισμούς για τους οποίους ο B έχει μεγαλύτερο SimBetUtil από τον A, τους κρατά αποθηκευμένους σε μια λίστα (request vector). Η λίστα αυτή αποστέλλεται στον A (message request). Τέλος, ο A, αφού λάβει τη λίστα request vector, προωθεί στον B τα αποθηκευμένα πακέτα του που έχουν προορισμό κάποιον από τους κόμβους που περιέχονται στο request vector. Σχηματικά, η ανταλλαγή πληροφορίας φαίνεται στο Σχήμα 2.2.

2.2. Ιδιωτικότητα σε δίκτυα

Παρατηρώντας τους αλγορίθμους δρομολόγησης βλέπουμε ότι κατά τη προώθηση δεδομένων, πάντα απαιτείται ανταλλαγή πληροφορίας μεταξύ των κόμβων, η οποία είναι ευαίσθητη. Κάθε αλγόριθμος, προκειμένου να αποφασίσει πως θα δρομολογήσει τα δεδομένα του, πρέπει να έχει γνώση τουλάχιστον του προορισμού των πακέτων που διαθέτει. Οι κόμβοι-χρήστες ενός δικτύου δεν θα ήθελαν η πληροφορία αυτή να μπορεί να αποκαλυφθεί, παραβιάζοντας έτσι την ιδιωτικότητά τους. Στις επόμενες παραγράφους ορίζουμε το πρόβλημα προστασίας της ιδιωτικότητας σε ένα δικτυακό περιβάλλον και τις προσεγγίσεις που έχουν προταθεί για την αντιμετώπισή του.

2.2.1. Ιδιωτικότητα σε ασύρματα κινητά δίκτυα

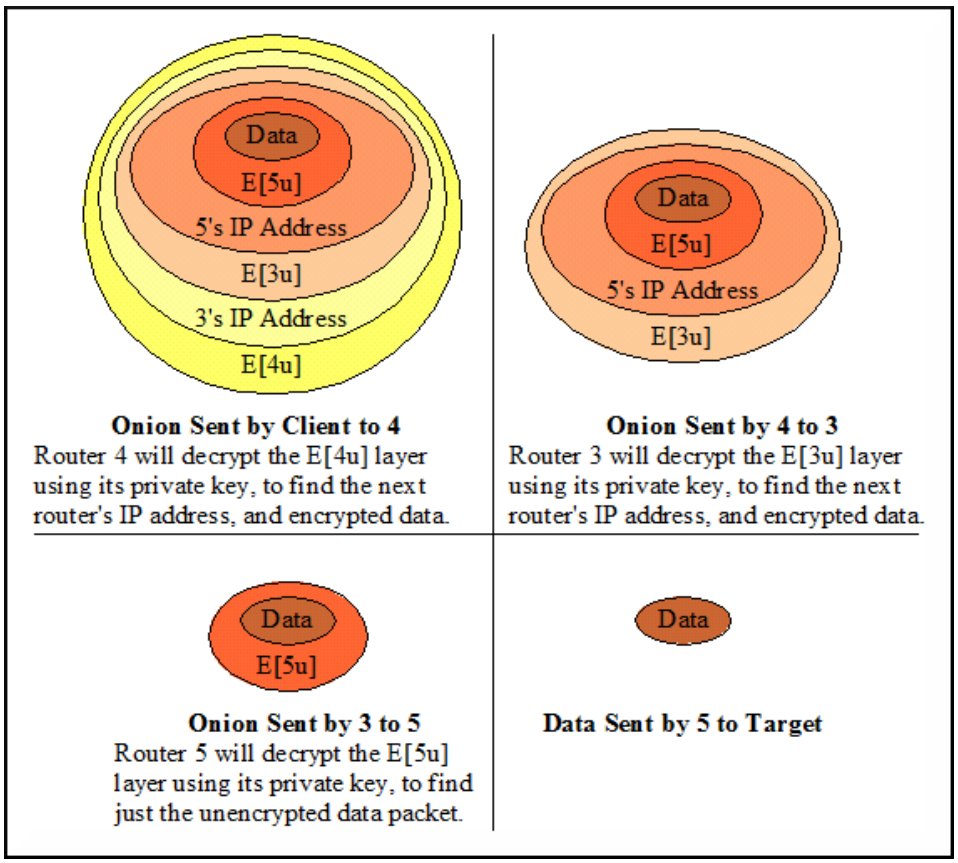
Όπως έχει αναφερθεί στο προηγούμενο κεφάλαιο, η ιδιωτικότητα αφορά την προστασία των ευαίσθητων προσωπικών πληροφοριών μιας οντότητας από μη εξουσιοδοτημένη προσπέλασή τους [14]. Η εξασφάλιση ιδιωτικότητας σε ένα δίκτυο ουσιαστικά ισοδυναμεί με την προστασία της ανωνυμίας των κόμβων κατά την

μεταξύ τους επικοινωνία. Για την προστασία της ιδιωτικότητας σε δίκτυα με σταθερή υποδομή έχουν προταθεί πολλές λύσεις από την επιστημονική κοινότητα, οι οποίες γενικώς δουλεύουν καλά στις περισσότερες περιπτώσεις. Ωστόσο, στα ασύρματα δίκτυα, τόσο η φύση του μέσου επικοινωνίας (αέρας), όσο και η κινητικότητα των κόμβων καθιστούν δύσκολη και τις περισσότερες φορές αδύνατη την εφαρμογή αυτών των μεθόδων. Η μετάδοση της πληροφορίας γίνεται με ευρεία εκπομπή (layer-2 broadcast) και κακόβουλοι χρήστες μπορούν να «κρυφακούνε» το κοινό μέσο και να εντοπίζει πακέτα ανεξάρτητα από το αν προορίζονται γι' αυτόν ή όχι (packet sniffing). Έτσι, ένας κακόβουλος χρήστης, ανεξάρτητα από το αν είναι ενδιάμεσος κόμβος σε κάποιο μονοπάτι μεταξύ αποστολέα και παραλήπτη, μπορεί να αποκτήσει την πληροφορία που απαιτείται για να σπάσει την ανωνυμία. Σε ένα περιβάλλον κινούμενων κόμβων, μάλιστα, όλοι οι κόμβοι του δικτύου μπορούν να λειτουργούν και ως δρομολογητές. Αυτό σημαίνει ότι η πληροφορία που υπάρχει σε ένα πακέτο (αποστολέας - παραλήπτης) μπορεί να γίνει διαθέσιμη σε οποιονδήποτε κόμβο του δικτύου και όχι μόνο στα access points. Το μοντέλο του δικτύου κινούμενων κόμβων, λοιπόν, κάνει το πρόβλημα ακόμη πιο δύσκολο.

Ο πιο κοινός τύπος επίθεσης κατά της ανωνυμίας σε ένα ασύρματο δίκτυο κινούμενων κόμβων είναι η ανάλυση της τηλεπικοινωνιακής κίνησης (traffic analysis). Ο λόγος που αποτελεί τον πιο κοινό τύπο επίθεσης είναι επειδή σε δίκτυα κινητών κόμβων κάθε κόμβος μπορεί να είναι και δρομολογητής. Η περίπτωση αυτή περιλαμβάνει ένα μεγάλο εύρος επιθέσεων που στηρίζονται στη γνώση που μπορεί να εξαχθεί από την παρατηρούμενη κίνηση στο δίκτυο. Η πιο απλή μέθοδος είναι η συλλογή πακέτων και η ανάλυση της πληροφορίας που περιέχεται στις κεφαλίδες (π.χ. Src, Dest, Next hop κ.λπ.). Σε αυτή την περίπτωση ο κακόβουλος χρήστης αποκαλύπτει άμεσα τον αποστολέα και τον παραλήπτη, «σπάζοντας» την ανωνυμία. Μια παραλλαγή αυτής της μεθόδου είναι το packet tracing. Σε αυτή τη επίθεση μπορεί να παρατηρηθεί η κίνηση ενός πακέτου καθώς διασχίζει το δίκτυο. Έτσι, χωρίς να απαιτείται γνώση του περιεχομένου ή των κεφαλίδων, μπορεί να αποκαλυφθούν τα άκρα της επικοινωνίας. Άλλη μια παραλλαγή αποτελούν οι επιθέσεις όπου οι εχθρικοί κόμβοι προσπαθούν να συσχετίσουν χρονικά τα πακέτα και να αναγνωρίσουν τη ροή της επικοινωνίας. Η πιο συνηθισμένη μορφή είναι η επίθεση TTL. Με βάση την τιμή του πεδίου TTL (time-to-live), ένας κακόβουλος

χρήστης μπορεί να υπολογίσει πόσο κοντά βρίσκεται στον αποστολέα ή στον προορισμό. Αυτές οι επιθέσεις είναι αποτελεσματικότερες αν οι εχθρικοί κόμβοι βρίσκονται κοντά σε ένα ή και στα δύο άκρα της επικοινωνίας.

Οι σχετικές εργασίες χρησιμοποιούν κάποιου είδους κρυπτογραφικό μηχανισμό για να υποστηρίξουν την ανωνυμία. Χαρακτηριστικό παράδειγμα αποτελεί η κρυπτογράφηση του μηνύματος σε επίπεδα (onion routing) [7]. Σε αυτό το μηχανισμό, το αρχικό πακέτο κρυπτογραφείται με το κλειδί κάθε ενδιάμεσου κόμβου-δρομολογητή. Αυτό απαιτεί εξ' αρχής καθορισμό του μονοπατιού. Κατά την προώθηση, κάθε ενδιάμεσος κόμβος αποκρυπτογραφεί το i-οστό επίπεδο του πακέτου και αποκαλύπτει μόνο τον επόμενο κόμβο που θα προωθήσει το πακέτο. Έτσι, οι ενδιάμεσοι κόμβοι δεν μπορούν να έχουν πλήρη εικόνα του μονοπατιού, ενώ τα δεδομένα και ο αποστολέας του πακέτου αποκαλύπτονται μόνο στο τελικό παραλήπτη (Σχήμα 2.3).



Σχήμα 2.3 Onion routing

Σε αυτήν την τεχνική εύκολα μπορεί να παρατηρήσει κανείς ότι το κόστος λόγω των πολλαπλών κρυπτογραφήσεων και αποκρυπτογραφήσεων είναι μεγάλο και ασύμφορο για ένα δίκτυο κινητών κόμβων. Επιπλέον, για να εφαρμοστεί η μέθοδος, όπως είπαμε, θα πρέπει ο αποστολέας να έχει καθορίσει εξ' αρχής τους ενδιάμεσους κόμβους που θα προωθήσουν το πακέτο. Αυτό συνεπάγεται ότι θα ξέρει κάποιο μονοπάτι προς τον τελικό προορισμό. Σε ένα δίκτυο κινητών κόμβων η ύπαρξη μονοπατιού, όπως έχουμε αναφέρει, δεν είναι πάντα εγγυημένη, συνεπώς, η μέθοδος αυτή είναι πρακτικά αδύνατο να εφαρμοστεί σε opportunistic δίκτυα.

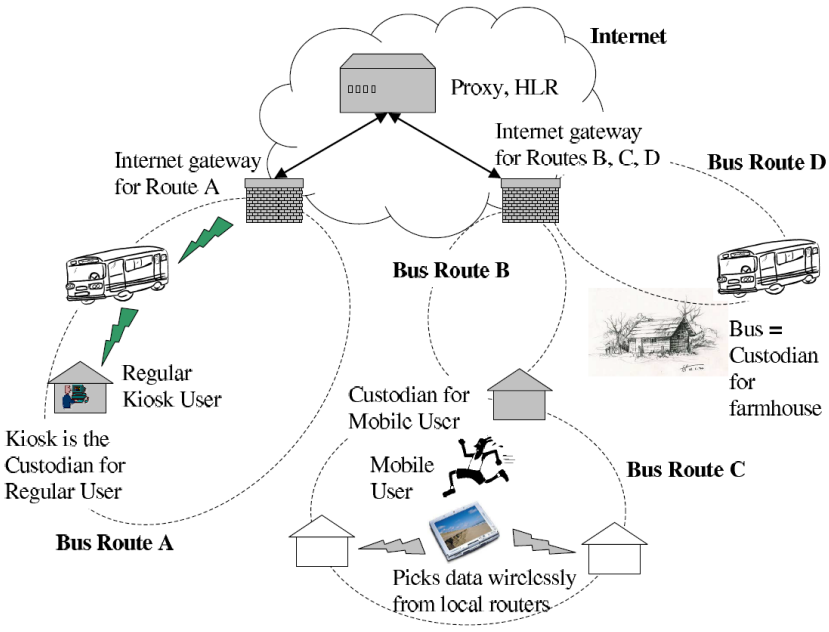
Μια προσέγγιση σε MANETs που προτείνεται από τον Choi [3] είναι η χρήση ψευδώνυμων, τα οποία αντικαθιστούν τις πραγματικές διευθύνσεις των κόμβων πάνω στα πακέτα και τα οποία αλλάζουν τακτικά. Τα ψευδώνυμα παράγονται από τους ίδιους τους κόμβους με βάση τα συμμετρικά κλειδιά που τους παρέχονται από μια αξιόπιστη οντότητα (trusted authority). Επίσης, προτείνεται η επιλογή ενός τυχαίου TTL και η ενσωμάτωσή του στο πραγματικό ώστε να αντιμετωπιστεί η επίθεση τύπου TTL. Η χρήση και διαχείριση των ψευδωνύμων δεν είναι εύκολο να εφαρμοστεί στην πράξη. Τα ψευδώνυμα αυτά πρέπει να αλλάζουν συχνά, αλλιώς η χρησιμότητά τους στο να μπερδεύουν τους κακόβουλους χρήστες παύει να ισχύει. Η αλλαγή των ψευδωνύμων εμπλέκει συνεχή επικοινωνία με την αξιόπιστη οντότητα, προκειμένου τα συμμετρικά κλειδιά να ανανεώνονται. Αυτό δεν μπορεί να είναι πάντα εφικτό, καθώς έχουμε να κάνουμε με ένα περιβάλλον κινούμενων κόμβων και μια τέτοιου είδους επικοινωνία δεν είναι εγγυημένη.

Μια άλλη λύση [20] στηρίζεται στη διατήρηση πολλαπλών μονοπατιών για κάθε παραλήπτη, ώστε η πληροφορία να διαχέεται στο δίκτυο σε πολλές κατευθύνσεις και να είναι δύσκολη η παρακολούθηση της ροής δεδομένων. Σε αυτή την περίπτωση η αδυναμία εφαρμογής αυτής της μεθόδου έγκειται στο γεγονός ότι δεν μπορούμε να έχουμε εξ' αρχής καθορισμό ενός μονοπατιού από τον αποστολέα στον παραλήπτη. Στη συγκεκριμένη περίπτωση μάλιστα, απαιτείται γνώση πολλών τέτοιων μονοπατιών, τα οποία δεν γνωρίζουμε αν υπάρχουν, λόγω της φύσης των δικτύων που μελετάμε (κινούμενοι κόμβοι).

2.2.2. Ιδιωτικότητα σε DTN/opportunistic δίκτυα

Στα DTN και στα opportunistic δίκτυα η προστασία της ιδιωτικότητας είναι ακόμη πιο δύσκολο έργο. Όπως έχουμε αναφέρει ήδη, η ύπαρξη μονοπατιού μεταξύ αποστολέα - παραλήπτη δεν είναι δεδομένη και η διαφορετική φιλοσοφία γύρω από τη δρομολόγηση δεν επιτρέπει τη χρήση των μέχρι τώρα αναφερθέντων τεχνικών. Στα δίκτυα αυτού του τύπου αποστολέας και προορισμός είναι πολύ πιθανόν να βρίσκονται σε διαφορετικές διαμερίσεις του δικτύου, άρα η χρήση παραδοσιακών τεχνικών προστασίας της ανωνυμίας αποτυγχάνει. Αυτό συμβαίνει επειδή οι τεχνικές που αναφέραμε στην προηγούμενη παράγραφο βασίζονται σε αλγόριθμους δρομολόγησης που προϋποθέτουν ότι ο αποστολέας και ο προορισμός να βρίσκονται στην ίδια διαμέριση. Για τους λόγους αυτούς οι εργασίες που έχουν γίνει είναι λιγοστές και οι μηχανισμοί κρυπτογράφησης/πιστοποίησης αποτελούν και εδώ το βασικό εργαλείο έρευνας.

Η πρώτη εργασία σχετική με προστασία της ανωνυμίας σε DTN προτάθηκε από τον Kate [9]. Το μοντέλο στο οποίο στηρίζεται είναι ένα DTN στο οποίο μπορεί να συμμετέχουν κόμβοι σε απομακρυσμένες περιοχές, όπως μια αγροτική περιοχή ή ένα χωριό, όπου δεν υπάρχει δυνατότητα μόνιμης σύνδεσης, λόγω γεωγραφικών περιορισμών (Σχήμα 2.4).

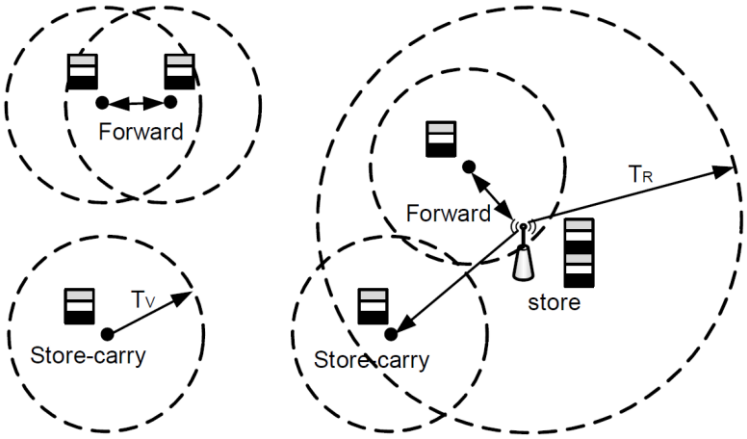


Σχήμα 2.4 Παράδειγμα DTN σε μια απομακρυσμένη περιοχή

Ο κρυπτογραφικός μηχανισμός χρησιμοποιεί την βασισμένη-στην-ταυτότητα κρυπτογράφηση (identity-based cryptography - IBE) [2]. Στην IBE θεωρούμε την ύπαρξη μιας έμπιστης οντότητας (public key generator - PKG) η οποία αναλαμβάνει τη δημιουργία και το διαμοιρασμό των αρχικών κλειδιών στους χρήστες. Σε ένα τέτοιο δίκτυο, το ρόλο των PKGs παίζουν κάποια σημεία υποδομής τα οποία ονομάζονται *kiosks*. Επιπλέον, υπάρχουν κινητοί κόμβοι που παίζουν το ρόλο δρομολογητών (mobile routers), όπως π.χ. ένα λεωφορείο και αναλαμβάνουν την προώθηση των δεδομένων στις δικτυακές πύλες (internet gateways), όπως φαίνεται στο Σχήμα 2.4. Μετά τον διαμοιρασμό των κλειδιών από τα PKGs, κάθε κόμβος διαθέτει δύο συμμετρικά κλειδιά, ένα για επικοινωνία και πιστοποίηση με κόμβους που ανήκουν στο ίδιο PKG (local key) και ένα για επικοινωνία με κόμβους εκτός του τοπικού PKG (long distance key). Με βάση τα κλειδιά αυτά οι χρήστες μπορούν να παράγουν μόνοι τους ψευδώνυμα τα οποία βασίζονται στα παραπάνω κλειδιά. Έτσι, οι πραγματικές ταυτότητες των χρηστών πάνω στα πακέτα αντικαθίστανται από τα ψευδώνυμα, εξασφαλίζοντας την ανωνυμία αποστολέα - παραλήπτη. Επιπλέον, επειδή τα ψευδώνυμα παράγονται με βάση τα κλειδιά των κόμβων, είναι δυνατή η πιστοποίηση κάθε κόμβου με βάση το ψευδώνυμό του. Κάθε φορά που ένας κόμβος θέλει να στείλει δεδομένα στο kiosk ή στον κινητό δρομολογητή, πιστοποιείται ότι το κλειδί του είναι έγκυρο χωρίς, ωστόσο, να απαιτείται αποκάλυψη της ταυτότητάς του. Η περαιτέρω δρομολόγηση του πακέτου από το kiosk ή το δρομολογητή μπορεί να γίνει με οποιοδήποτε πρωτόκολλο δρομολόγησης. Αντίστοιχα, ο μηχανισμός πιστοποίησης εφαρμόζεται και στον παραλήπτη για την πιστοποίηση της εγκυρότητας του ψευδώνυμού του.

Τα βασικότερο μειονέκτημα αυτής της προσέγγισης είναι ότι ουσιαστικά εστιάζει στο να παρέχει ανωνυμία μόνο κατά το πρώτο και το τελευταίο άλμα της επικοινωνίας. Λεπτομέρειες για το πως εξασφαλίζεται η ανώνυμη προώθηση των πακέτων στα ενδιάμεσα άλματα της επικοινωνίας δεν παρέχονται. Επίσης, το μοντέλο στο οποίο αναπτύσσεται η προσέγγιση είναι πολύ εξειδικευμένο και αποκλείει ή κάνει πολύ δύσκολη την προσαρμογή της μεθόδου σε περιπτώσεις όπου η υποδομή απουσιάζει. Γενικότερα, μπορούμε να πούμε ότι πρόκειται για μια πολύ εξειδικευμένη προσέγγιση που προσφέρει υποτυπώδη μόνο ανωνυμία.

Ένα άλλο γνωστό πρωτόκολλο είναι το SPRING [13]. Το μοντέλο του δικτύου που χρησιμοποιείται είναι ένα DTN, όπου οι κόμβοι είναι οχήματα κινούμενα στο περιβάλλον μιας πόλης (vehicular DTN).



Σχήμα 2.5 Παράδειγμα ενός vehicular DTN

Επιπλέον, στο δίκτυο υπάρχουν σταθεροί κόμβοι (roadside units-RSUs) όπως φαίνεται στο Σχήμα 2.5 οι οποίοι δρουν βοηθητικά, διευκολύνοντας την επικοινωνία μεταξύ των οχημάτων. Τα RSUs τοποθετούνται στρατηγικά σε περιοχές με μεγάλη τηλεπικοινωνιακή κίνηση, π.χ. διασταυρώσεις. Επιπλέον, γίνεται η υπόθεση ότι τα RSUs είναι αξιόπιστα και δεν πρόκειται να προσπαθήσουν να «σπάσουν» την ανωνυμία. Κατά την προώθηση των πακέτων χρησιμοποιείται ένας μηχανισμός πιστοποίησης, ο οποίος βασίζεται στην τεχνική *CPPA* (Conditional Privacy Preserving Authentication) [12]. Η παραπάνω τεχνική εξασφαλίζει ανώνυμη πιστοποίηση μεταξύ των κόμβων με χρήση ασύμμετρης κρυπτογράφησης, χωρίς να αποκαλύπτεται πληροφορία σχετική με την ταυτότητά τους. Τα κρυπτογραφημένα πακέτα προωθούνται μόνο ανάμεσα σε πιστοποιημένους κόμβους και αντί των πραγματικών διευθύνσεων, χρησιμοποιείται η γεωγραφική θέση του κόμβου. Όταν ένας κόμβος, λόγω περιορισμών μνήμης, δεν μπορεί πλέον να αποθηκεύσει άλλα μηνύματα, τότε προσπαθεί να τα προωθήσει, κατά προτεραιότητα, προς κάποιο κοντινό RSU. Ο αποθηκευτικός χώρος ενός RSU είναι πολύ μεγαλύτερος από έναν κόμβο-όχημα και επιπλέον, εγγυάται την προστασία του μηνύματος από κακόβουλους χρήστες. Το RSU αφού πιστοποιήσει τον αποστολέα και λάβει το πακέτο, θα το επανακρυπτογραφήσει, ώστε σε επόμενη προώθησή του να μην μπορεί

να αναγνωριστεί, παρά μόνο από τον τελικό παραλήπτη. Έτσι, αποφεύγεται η επίθεση τύπου packet tracing. Αν δεν υπάρχει δυνατότητα προώθησης του πακέτου σε κάποιο κοντινό RSU, τότε επιλέγεται κάποιο διερχόμενο όχημα.

Η παραπάνω προσέγγιση έχει τρεις βασικές αδυναμίες. Πρώτον, γίνεται χρήση ασύμμετρης κρυπτογράφησης κάτι που θέλουμε να αποφεύγουμε στο είδος των δικτύων που μελετάμε. Όπως έχουμε αναφέρει, οι κόμβοι είναι φορητές συσκευές με περιορισμένες υπολογιστικές δυνατότητες και τέτοιου είδους λειτουργίες είναι καλό να αποφεύγονται. Δεύτερον, η αντικατάσταση του προορισμού με γεωγραφικό προσδιορισμό δεν μας εξασφαλίζει ανωνυμία. Εάν στην περιοχή που ορίζεται ως προορισμός του πακέτου το μόνο όχημα που υπάρχει είναι ο παραλήπτης, τότε εύκολα ο κόμβος που θα του προωθήσει τελευταίος το πακέτο (αν πρόκειται για άλλο όχημα) θα καταλάβει ότι αυτός είναι ο προορισμός. Τέλος, η χρήση αξιόπιστης υποδομής στο μοντέλο διευκολύνει τη λύση του προβλήματος. Έχοντας κάποιους σταθερούς αξιόπιστους κόμβους και προωθώντας τα πακέτα κατά προτεραιότητα σε αυτούς κάνει το πρόβλημα πιο εύκολο. Αυτό ισχύει αφού μεταξύ αποστολέα και παραλήπτη, με μεγάλη πιθανότητα, θα παρεμβάλλονται κόμβοι αξιόπιστοι που δεν θα προσπαθήσουν να παραβιάσουν την ανωνυμία.

Οι λύσεις που έχουμε αναφέρει ως τώρα βασίζονται σε μηχανισμούς κρυπτογράφησης, συχνά περίπλοκους και υπολογιστικά ακριβούς για τους κινητούς κόμβους. Μια διαφορετική προσέγγιση για την προστασία της ανωνυμίας παρουσιάζεται στο [15]. Εδώ οι συγγραφείς προτείνουν τη χρήση φίλτρων Bloom για την αναπαράσταση της πληροφορίας δρομολόγησης. Η δρομολόγηση στηρίζεται στη ύπαρξη κοινωνικών σχέσεων μεταξύ των κόμβων. Κάθε κόμβος έχει μια λίστα «φίλων» οι οποία αναπαριστά τους κόμβους με τους οποίους σχετίζεται στενά. Κάθε κόμβος-αποστολέας επιλέγει εξ' αρχής τους κόμβους που θα μπορούν να προωθούν το προς αποστολή πακέτο (forwarders list). Η επιλογή των κόμβων αυτών από τον αποστολέα γίνεται με βάση τη κοινωνική σχέση του με αυτούς (επιλογή από τη λίστα «φίλων»). Η προώθηση του πακέτου γίνεται μεταξύ κόμβων που ανήκουν στην παραπάνω λίστα. Για τη διατήρηση της ιδιωτικότητας απαιτείται η λίστα των forwarders να μην αποκαλύπτεται ποτέ πλήρως. Για το λόγο αυτό προτείνεται η τροποποίηση της λίστας των forwarders και η αναπαράστασή της με χρήση φίλτρων

Bloom. Με τον τρόπο αυτό, όταν δύο κόμβοι συναντιούνται, ο ένας κόμβος μπορεί να ελέγξει αν ο άλλος βρίσκεται στη λίστα των forwarders (membership check), ώστε να του προωθήσει το πακέτο, χωρίς ωστόσο, να έχει πρόσβαση σε όλη την πληροφορία της λίστας.

Η παραπάνω ιδέα είναι απλή και δεν επιβαρύνεται από το κόστος της κρυπτογράφησης. Από την άλλη, όμως, η αποτελεσματικότητά της όσον αφορά την προστασία της ανωνυμίας δεν είναι καθόλου εγγυημένη. Το φίλτρο που αναπαριστά τη λίστα των forwarders δημιουργείται με βάση τις πραγματικές διευθύνσεις των κόμβων, οι οποίες είναι γνωστές σε όλους τους υπόλοιπους κόμβους του δικτύου. Κάποιος κακόβουλος κόμβος μπορεί πολύ εύκολα να σπάσει την ανωνυμία ελέγχοντας για όλους τους κόμβους του δικτύου αν υπάρχουν στη λίστα (βίαιη επίθεση - brute force). Ο παραπάνω έλεγχος δεν είναι δύσκολο να γίνει ακόμα και για μεγάλο πλήθος κόμβων, αφού ο έλεγχος ύπαρξης ενός στοιχείου στο φίλτρο περιλαμβάνει απλούς υπολογισμούς συναρτήσεων κατακερματισμού Με αυτόν τον τρόπο μπορεί, με μεγάλη πιθανότητα, να αποκαλύψει ποιοι κόμβοι βρίσκονται στη λίστα, αποκαλύπτοντας την πληροφορία δρομολόγησης που θέλουμε να αποκρύψουμε.

ΚΕΦΑΛΑΙΟ 3. ΠΡΟΤΕΙΝΟΜΕΝΟΣ ΑΛΓΟΡΙΘΜΟΣ

Στο κεφάλαιο αυτό περιγράφουμε αναλυτικά το πρόβλημα της προστασίας της ιδιωτικότητας και τις δυσκολίες που παρουσιάζει η αντιμετώπισή του. Έπειτα παρουσιάζουμε τον αλγόριθμο που προτείνουμε για την προστασία της ανωνυμίας σε opportunistic δίκτυα. Η λύση μας βασίστηκε στο γνωστό αλγορίθμου SimBet που περιγράψαμε σε προηγούμενη ενότητα. Η ιδέα βασίζεται στη χρήση φίλτρων Bloom (βλέπε Παράρτημα). Τα φίλτρα Bloom είναι μια δομή δεδομένων η οποία μπορεί να χρησιμοποιηθεί για την αναπαράσταση στοιχείων ενός συνόλου με μορφή δυαδικού πίνακα. Κάθε στοιχείο που πρόκειται να εισαχθεί στο φίλτρο δίνεται ως είσοδος σε k συναρτήσεις κατακερματισμού. Οι τιμές που επιστρέφουν οι συναρτήσεις αποτελούν τις θέσεις που θα τεθούν 1 στον δυαδικό πίνακα, δίνοντας έτσι μια δυαδική αναπαράσταση του στοιχείου. Δοθείσης αυτής της μορφής δεν μπορούμε να εξάγουμε το σύνολο των στοιχείων που αναπαριστά. Η μόνη λειτουργία που μπορούμε να εφαρμόσουμε είναι η επερώτηση για την ύπαρξη ενός στοιχείου. Η παραπάνω ιδιότητα μας επιτρέπει να χρησιμοποιήσουμε τα φίλτρα Bloom για την αναπαράσταση των διευθύνσεων των κόμβων καθώς και των επαφών που ανταλλάσσονται κατά τη δρομολόγηση στον SimBet. Ο προτεινόμενος αλγόριθμος

παρέχει προστασία της ανωνυμίας κατά την επικοινωνία, διατηρώντας παράλληλα την αποτελεσματικότητα του αρχικού αλγόριθμου σε πολύ καλά επίπεδα.

3.1. Περιγραφή του προβλήματος

Με βάση όσα αναφέραμε στο προηγούμενο κεφάλαιο γίνεται σαφές ότι η εξασφάλιση ανώνυμης επικοινωνίας σε ένα δίκτυο κινούμενων κόμβων είναι ένα δύσκολο ερευνητικό πρόβλημα. Η απόκρυψη των άκρων της επικοινωνίας δεν είναι εύκολο να επιτευχθεί καθώς η διεύθυνση του αποστολέα και του προορισμού ενός πακέτου περιέχεται στις αντίστοιχες κεφαλίδες του. Γνωρίζοντας τον αποστολέα και τον προορισμό, άμεσα αποκαλύπτονται τα άκρα της επικοινωνίας. Συνεπώς, οι ενδιάμεσοι κόμβοι που αναλαμβάνουν να προωθήσουν τα πακέτα έχουν πλήρη πρόσβαση σε όλη αυτή την πληροφορία και μπορούν να σπάσουν άμεσα την ανώνυμη επικοινωνία. Η απόκρυψη του αποστολέα σε ένα πακέτο είναι δυνατή, αφού η διαδικασία της δρομολόγησης δεν σχετίζεται πάντα με αυτή την πληροφορία. Ωστόσο, για να αποφασίσουμε την περαιτέρω προώθηση ενός πακέτου πρέπει απαραίτητα να ξέρουμε τον προορισμό του. Αυτό ισχύει σε όλα τα πρωτόκολλα επικοινωνίας, ανεξαρτήτως δικτυακού μοντέλου ή φυσικού μέσου και είναι ο σημαντικότερος παράγοντας δυσκολίας στο πρόβλημα που μελετάμε. Ένα παράδειγμα όπου φαίνεται η παραπάνω συσχέτιση αποτελεί ο υπολογισμός μετρικών οι οποίες καθορίζουν ποιος κόμβος θα προωθήσει ένα πακέτο. Οι μετρικές που χρησιμοποιούνται στους αλγορίθμους δρομολόγησης, που είδαμε στο κεφάλαιο 2, αναπαριστούν την εγγύτητα ενός κόμβου στον προορισμό, είτε αυτό εκφράζεται με κάποια πιθανότητα, είτε με πλήθος αλμάτων κ.λπ. Γενικεύοντας, λοιπόν, μπορούμε να πούμε ότι το επόμενο άλμα στην επικοινωνία εκφράζεται πάντα ως συνάρτηση του τελικού προορισμού. Συνεπώς από τη φύση του προβλήματος, βλέπουμε ότι η λειτουργία της δρομολόγησης και η ανώνυμη επικοινωνία δεν συμβαδίζουν.

Οι συνήθεις τακτικές που ακολουθούνται για την προστασία της ανωνυμίας στηρίζονται, όπως είδαμε, κυρίως στην κρυπτογράφηση και στην χρήση ψευδωνύμων. Οι λύσεις αυτές είτε είναι υπολογιστικά ακριβές για ένα ασύρματο δίκτυο, είτε προϋποθέτουν την ύπαρξη πλήρους μονοπατιού μεταξύ αποστολέα-παραλήπτη. Στην πρώτη περίπτωση το υπολογιστικό κόστος είναι υψηλό μιας και οι

κόμβοι είναι φορητές συσκευές με περιορισμένες υπολογιστικές δυνατότητες, ενώ στη δεύτερη, η ύπαρξη μονοπατιού δεν είναι εγγυημένη. Στα opportunistic δίκτυα τέτοιες προσεγγίσεις είναι αδύνατο να εφαρμοστούν, καθώς δεν μπορούμε να έχουμε γνώση της τοπολογίας του δικτύου, πέρα από την τοπική γειτονιά. Συνεπώς, δεν μπορεί να καθοριστεί μέσω ποιων κόμβων θα δρομολογηθούν τα πακέτα στον τελικό προορισμό, ώστε τα άκρα της επικοινωνίας και οι ενδιάμεσοι κόμβοι να αναπαρασταθούν με κάποιες συμβολικές διευθύνσεις (ψευδώνυμα). Επιπλέον, η χρήση υποδομής που υποθέτουν κάποιες εργασίες μπορεί να διευκολύνει την επίλυση του προβλήματος, αλλά η ύπαρξή της δεν μπορεί να είναι πάντα εγγυημένη (π.χ. περιπτώσεις φυσικών καταστροφών).

Στην κατηγορία των αλγορίθμων δρομολόγησης με αρχές κοινωνικής δικτύωσης, οι κίνδυνοι είναι ακόμη μεγαλύτεροι. Εδώ, ο υπολογισμός των μετρικών συχνά δεν απαιτεί τη γνώση μόνο του προορισμού, αλλά και επιπλέον πληροφορία. Στον SimBet για παράδειγμα, όπως περιγράψαμε στο Κεφ. 2, όταν δύο κόμβοι εκτελούν το πρωτόκολλο, για να γίνει ο υπολογισμός των μετρικών betweenness και similarity πρέπει να γίνει ανταλλαγή των επαφών των δύο κόμβων. Οι επαφές ενός κόμβου αποτελούν ευαίσθητη πληροφορία για έναν κόμβο και η αποκάλυψή τους παραβιάζει την ιδιωτικότητα. Συνεπώς, έκτος από την πληροφορία που περιέχεται στα πακέτα και η οποία μπορεί άμεσα να παραβιάσει την ανωνυμία, πρέπει να προστατεύσουμε και τις επαφές που ανταλλάσσονται. Επίσης, η τακτική που ακολουθείται στα opportunistic δίκτυα (store, carry and forward), από τη φύση της αποκαλύπτει τον τελικό προορισμό στο τελευταίο άλμα της επικοινωνίας. Ο κόμβος που θα παραδώσει το πακέτο στον τελικό παραλήπτη, αναπόφευκτα πρέπει να ξέρει τη διεύθυνσή του. Αυτό οδηγεί σε άμεση παραβίαση της ανωνυμίας του παραλήπτη.

Από τα παραπάνω συμπεραίνουμε ότι η εξασφάλιση ανωνυμίας σε opportunistic δίκτυα αντιμετωπίζει δυσκολίες που αφορούν τόσο στη φύση του προβλήματος, όσο και στις ιδιαίτερες συνθήκες που παρουσιάζει αυτό το μοντέλο δικτύου. Στη συνέχεια, παρουσιάζουμε την προσέγγιση μας πάνω στο πρόβλημα η οποία εξασφαλίζει προστασία της ανωνυμίας, με μικρότερο υπολογιστικό κόστος από τις υπάρχουσες μεθόδους. Το βασικό εργαλείο που χρησιμοποιήσαμε είναι τα φίλτρα Bloom, τα οποία μας προσφέρουν μια αναπαράσταση της πληροφορίας που

απαιτείται για την επικοινωνία στο δίκτυο με μορφή δυαδικού πίνακα. Η μορφή αυτή μας επιτρέπει να υπολογίζουμε με τις μετρικές ενός κόμβου για έναν προορισμό χωρίς να γνωρίζουμε ποια είναι η ταυτότητα του προορισμού. Ο υπολογισμός αυτός εμπεριέχει μια πιθανότητα σφάλματος, η οποία οφείλεται στη φύση της ίδιας της δομής των φίλτρων, χωρίς ωστόσο, αυτό να επηρεάζει σημαντικά την απόδοση του νέου αλγόριθμου όπως θα φανεί κατά την αξιολόγησή του.

3.2. Χρήση φίλτρων Bloom στη δρομολόγηση

Τα φίλτρα Bloom αποτελούν μια ειδική δομή δεδομένων η οποία μας επιτρέπει να αναπαραστήσουμε ένα σύνολο στοιχείων με χρήση ενός μονοδιάστατου δυαδικού πίνακα. Η αναπαράσταση αυτή είναι ιδιαίτερα χρήσιμη, αφού δοθέντος του δυαδικού πίνακα, δεν μπορεί να εξαχθεί άμεσα από αυτόν το σύνολο των στοιχείων που αναπαριστά. Η μόνη λειτουργία που μπορεί να εφαρμοστεί σε ένα φίλτρο Bloom για την εξαγωγή της πληροφορίας που περιέχει, είναι η επερώτηση για την ύπαρξη ενός στοιχείου. Αυτή η ιδιότητα μας επιτρέπει να «κρύψουμε» την πληροφορία που διακινείται στη δρομολόγηση, αναπαριστώντας την με ένα φίλτρο Bloom. Στην περίπτωση μας τα στοιχεία που επιλέγουμε να αναπαραστήσουμε με ένα φίλτρο είναι *οι επαφές ενός κόμβου και οι προορισμοί των πακέτων του*. Όλη λοιπόν η ευαίσθητη πληροφορία που απαιτείται για τη λειτουργία του αλγορίθμου αναπαρίσταται με ένα φίλτρο Bloom. Οι κόμβοι που επικοινωνούν μπορούν να ελέγχουν αν μια επαφή περιέχεται στο φίλτρο, αλλά δεν μπορούν να έχουν πρόσβαση σε όλη τη λίστα των επαφών. Αντίστοιχα, μπορούν να υπολογίζουν με μεγάλη ακρίβεια τις μετρικές που απαιτούνται για τον προορισμό ενός πακέτου χωρίς να έχουν γνώση της πραγματικής του διεύθυνσης.

Η χρήση φίλτρων Bloom με τον παραπάνω τρόπο προϋποθέτει ότι κάθε κόμβος δεν θα χρησιμοποιεί πλέον την πραγματική του διεύθυνση. Η χρήση των πραγματικών διευθύνσεων για την αναπαράσταση των επαφών δεν ενδείκνυται, αφού κάποιος κακόβουλος θα μπορούσε εύκολα να αναγνωρίσει τον προορισμό ενός πακέτου από τις κεφαλίδες του. Επίσης, θα μπορούσε να ελέγξει τη ύπαρξη κόμβων στο φίλτρο, δοκιμάζοντας τις πραγματικές διευθύνσεις (οι οποίες θα ήταν γνωστές) για όλους τους κόμβους του δικτύου (brute force). Έτσι, με μεγάλη πιθανότητα, θα αποκάλυπτε

το σύνολο των επαφών που περιέχονται στο φίλτρο. Συνεπώς, η χρήση φίλτρων Bloom απαιτεί τον ορισμό ενός νέου τρόπου αναπαράστασης των διευθύνσεων των κόμβων, έτσι ώστε το πρωτόκολλο να παρέχει ανωνυμία, αλλά ταυτόχρονα να μπορεί να λειτουργεί σωστά.

Ο υπολογισμός των μετρικών με χρήση φίλτρων Bloom γίνεται έχοντας ένα περιθώριο λάθους, το οποίο οφείλεται στην κατασκευή των φίλτρων. Το περιθώριο αυτό σχετίζεται με την πιθανότητα να έχουμε false positives, δηλ. μια ενδεχόμενη θετική απάντηση σχετικά με την ύπαρξη ενός στοιχείου, να είναι λάθος. Η ύπαρξη false positives θα μπορούσε να είναι ένας ανασταλτικός παράγοντας για τη χρήση φίλτρων Bloom. Όμως, όπως θα δούμε, δεν επηρεάζει σημαντικά την απόδοση και τη σωστή λειτουργία του αλγορίθμου.

3.3. Ο αλγόριθμος SimBet-BF

Υιοθετώντας την παραπάνω προσέγγιση και βασισμένοι στον αλγόριθμο SimBet, διατυπώνουμε εν συνεχεία τη λύση που προτείνουμε για εξασφάλιση ανώνυμης επικοινωνίας σε opportunistic δίκτυα.

3.3.1. Αναπαράσταση κόμβων και επαφών με χρήση φίλτρων Bloom

Όπως αναφέραμε, οι διευθύνσεις των κόμβων δεν μπορούν να αναπαρίστανται με την πραγματική τους μορφή πάνω στο πακέτο, γιατί αυτό θα οδηγούσε σε σπάσιμο της ανωνυμίας. Συνεπώς, απαιτείται μια διαφορετική μορφή αναπαράστασης η οποία δεν θα επιτρέπει τη συσχέτιση με την πραγματική διεύθυνση, αλλά ταυτόχρονα θα είναι χρηστική και δεν θα επηρεάζει τη σωστή λειτουργία του αλγορίθμου.

Για να επιτύχουμε μια τέτοια αναπαράσταση θεωρούμε την ύπαρξη μιας αξιόπιστης οντότητας (trusted authority - TA[4]) η οποία αναλαμβάνει τη δημιουργία κλειδιών τα οποία αντιστοιχίζονται στα ζεύγη των κόμβων του δικτύου. Η παραδοχή αυτή είναι συνήθης σε παρόμοιες εργασίες. Τα κλειδιά αυτά παράγονται πριν την έναρξη

[4] Στο υπόλοιπο της διατριβής η αξιόπιστη οντότητα θα αναφέρεται με το ακρωνύμιο TA

λειτουργίας του δικτύου (off-line). Για κάθε ζεύγος επικοινωνίας απαιτούμε την ύπαρξη 2 κλειδιών, ένα για κάθε κατεύθυνση επικοινωνίας. Στο σημείο αυτό θα μπορούσε κανείς να αναρωτηθεί γιατί 2 κλειδιά ανά ζεύγος επικοινωνίας; Η χρήση συμμετρικών κλειδιών στις παραδοσιακές μεθόδους, απαιτεί ένα κλειδί ανά ζεύγος αποστολέα-παραλήπτη. Ωστόσο, η δημιουργία 2 κλειδιών είναι απαραίτητη για να ξεχωρίζουμε τη ροή της επικοινωνίας μεταξύ δύο κόμβων (ποιος από τους δύο είναι ο αποστολέας και ποιος ο παραλήπτης). Η γνώση αυτή είναι απαραίτητη γιατί επηρεάζει τον υπολογισμό της μετρικής similarity, όπως θα δούμε παρακάτω. Άρα, λοιπόν, για ένα ζεύγος επικοινωνίας, έστω A, B, παράγονται τα κλειδιά K_{AB} και K_{BA}. Γενικά, με K_{ij} συμβολίζουμε το κλειδί που αναφέρεται στη ροή επικοινωνίας από τον i στον j ενώ με K_{ji} συμβολίζουμε το κλειδί που αναφέρεται στη ροή επικοινωνίας από τον j στον i. Τα i και j συμβολίζουν τις πραγματικές διευθύνσεις των κόμβων, έτσι όπως αυτές γίνονται γνωστές στους υπόλοιπους κόμβους του δικτύου. Επιπλέον, η ΤΑ παράγει και κάποια κλειδιά τα οποία αναφέρονται σε *εικονικούς κόμβους* με σκοπό να δυσκολέψει το έργο των κακόβουλων χρηστών[5]. Τα κλειδιά αυτά θα τα αναφέρουμε ως *εικονικά κλειδιά*. Τα κλειδιά αυτά, λοιπόν, εικονικά και μη, δεν παρέχονται στους κόμβους με αυτή τη μορφή, αλλά το καθένα εισάγεται σε ένα φίλτρο Bloom. Συνεπώς, σε έναν κόμβο i η ΤΑ παρέχει για κάθε κόμβο-προορισμό j τη διεύθυνση:

$$j \leftrightarrow dest_i^j = BF(K_{ij}), \forall j \qquad \text{Εξ 3.1}$$

Αυτή είναι η διεύθυνση που θέτει ως προορισμό ο κόμβος i πάνω στα πακέτα που θέλει να στείλει στον j. Κάθε κλειδί, λοιπόν, που αναφέρεται σε έναν πιθανό προορισμό, άσχετα από το αν είναι έγκυρος ή όχι εισάγεται σε ένα φίλτρο Bloom και παρέχεται στον i. Αντίστοιχα, για την αναγνώριση του αποστολέα j ενός πακέτου που προορίζεται για τον i, η ΤΑ παρέχει στον i τις διευθύνσεις:

$$j \leftrightarrow src_i^j = BF(K_{ji}), \forall j \qquad \text{Εξ 3.2}$$

[5] Ο λόγος δημιουργίας και χρήσης των κλειδιών αυτών θα αναλυθεί στην ενότητα 3.4

Αυτή είναι η διεύθυνση που θα περιέχεται στα πακέτα που λαμβάνει ο i από τον j. Σε αυτά τα φίλτρα, δηλαδή, περιέχονται τα κλειδιά που σχετίζονται με την αντίστροφη ροή πληροφορίας.

Εκτός από τα παραπάνω κλειδιά, για κάθε κόμβο i παράγεται ένα ατομικό κλειδί, K_i, το οποίο χρησιμεύει, όπως θα δούμε στη συνέχεια, στο να αποτρέψει κακόβουλους χρήστες να χρησιμοποιούν μη έγκυρες διευθύνσεις.

Συνολικά, λοιπόν, αν N είναι το πλήθος των πραγματικών κόμβων στο δίκτυο και N* το πλήθος των εικονικών κόμβων, παράγονται $2\times(N+N^*)\times(N+N^*-1)/2 + N + N^* = (N+N^*)^2$ κλειδιά. Τα παραπάνω κλειδιά, χρησιμοποιούνται για να δημιουργήσουμε τη νέα αναπαράσταση των διευθύνσεων των κόμβων. Κάθε κόμβος i του δικτύου έχει πλέον στη διάθεσή του N+N*-1 διευθύνσεις της μορφής:

$$ip_v \leftrightarrow id_i^v = BF(\sum_{j\neq i,v} K_{ji} + K_v), \forall v \qquad \text{Εξ 3.3}$$

Το φίλτρο αυτό περιέχει ως στοιχεία τα κλειδιά K_{ji} και το K_v. Με K_{ji} συμβολίζουμε το κλειδί του ζεύγους j, i που αναφέρεται στη ροή επικοινωνίας από τον j στον i και K_v το ατομικό κλειδί του κόμβου v. Με ip_v συμβολίζουμε τη διεύθυνση που παρέχει η ΤΑ στον i για την αναγνώριση του v. Η διεύθυνση αυτή δεν είναι η πραγματική διεύθυνση του v και επιπλέον, μέσω αυτής δεν μπορεί να υπάρξει κάποιος συσχετισμός με την πραγματική. Το φίλτρο id_i^v αποτελεί τη διεύθυνση με την οποία ένας κόμβος i γίνεται γνωστός στον κόμβο v. Όπως, βλέπουμε, είναι φτιαγμένο με τέτοιο τρόπο, ώστε να περιέχει όλα τα έγκυρα κλειδιά που αναφέρονται στον i ως προορισμό. Αυτός ο τρόπος αναπαράστασης εξασφαλίζει το σωστό υπολογισμό των μετρικών, κυρίως του similarity, όπως θα δούμε στην επόμενη ενότητα. Επίσης, βλέπουμε ότι για κάθε διεύθυνση της μορφής id_i^v σκόπιμα δεν περιλαμβάνουμε το κλειδί K_{vi}. Η εισαγωγή του κλειδιού αυτού θα οδηγούσε σε άμεση αναγνώριση του i από τον v. Ο v θα μπορούσε, να κάνει δοκιμές υπολογίζοντας το λογικό AND του id_i^v και του $dest_i^j$ για κάθε j. Για το $dest_i^v$ το αποτέλεσμα θα ήταν το ίδιο το $dest_i^v$ (είναι το μοναδικό j για το οποίο θα ισχύει), οπότε ο v θα μπορούσε να συμπεράνει με βεβαιότητα ότι ο κόμβος που του έστειλε το id_i^v είναι ο i. Επιπλέον, είναι απαραίτητο

να μην χρησιμοποιήσουμε κάποια από τα εικονικά κλειδιά κατά την κατασκευή των φίλτρων-διευθύνσεων. Όπως αναφέραμε, τα κλειδιά που αντιστοιχούν σε έγκυρους κόμβους είναι απαραίτητα για το σωστό υπολογισμό των μετρικών, άρα πρέπει να περιέχονται στα έγκυρα id. Αντίθετα, η απουσία τους στις εικονικές διευθύνσεις δεν θα προκαλέσει κανένα λάθος υπολογισμό, αφού δεν πρόκειται ποτέ να χρησιμοποιηθούν. Το να μην συμπεριλάβουμε έναν σταθερό αριθμό εικονικών κλειδιών, τα οποία επιλέγονται τυχαία από την ΤΑ, προστατεύει την ανωνυμία των κόμβων κατά την άμεση επικοινωνία[6]. Συνεπώς, η ΤΑ θα φροντίσει να κατασκευάσει τα φίλτρα διευθύνσεων με τέτοιο τρόπο, ώστε να αφαιρούνται εικονικά κλειδιά από τα έγκυρα id και οποιαδήποτε κλειδιά από τα εικονικά id. Το πλήθος των εικονικών κλειδιών που αφαιρούνται θα το συμβολίζουμε με x.

Ουσιαστικά, λοιπόν, κάθε κόμβος χρησιμοποιεί διαφορετική ταυτότητα για την αναγνώρισή του από κάθε άλλο κόμβο v του δικτύου. Οι παραπάνω διευθύνσεις παρέχονται στον κόμβο i από την ΤΑ που παρήγαγε τα κλειδιά, οπότε ο κόμβος i δεν γνωρίζει τα αρχικά κλειδιά τα οποία περιέχονται στο φίλτρο και δεν μπορεί να έχει γνώση σχετικά με την κατασκευή του. Επιπλέον, δεν υπάρχει τρόπος να ξεχωρίσει ποιες από τις διευθύνσεις που του παρέχονται είναι προς έγκυρους κόμβους και ποιες προς εικονικούς.

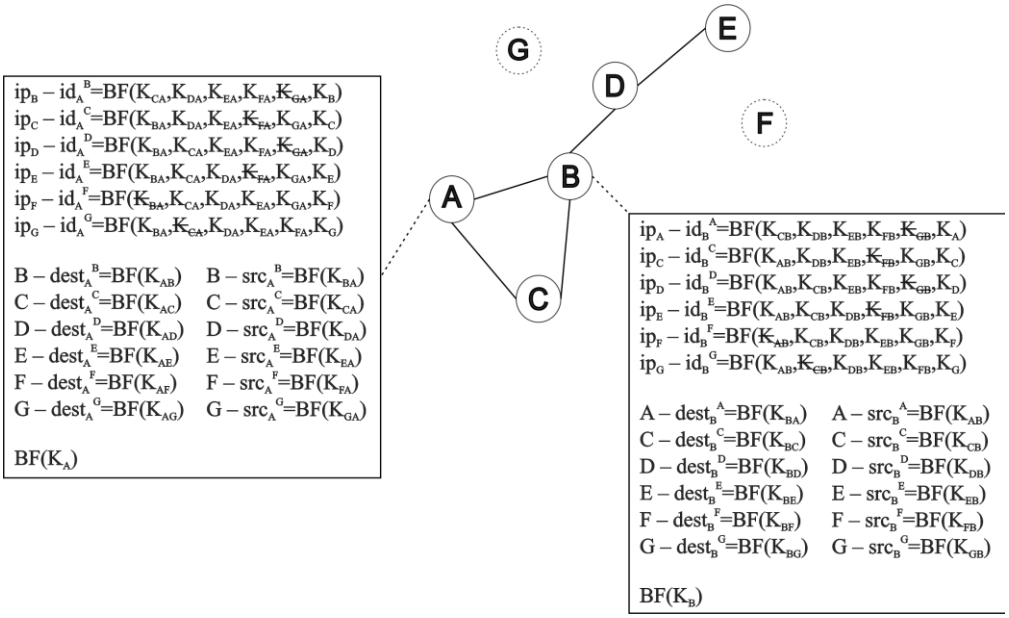
Τέλος, η ΤΑ παρέχει στον κόμβο i και το φίλτρο Bloom που περιέχει το ατομικό του κλειδί K_i, το $BF(K_i)$. Το φίλτρο αυτό είναι ιδιαίτερα χρήσιμο, καθώς επιτρέπει στον κόμβο i να ελέγχει αν οι διευθύνσεις οι οποίες λαμβάνει από τους κόμβους με τους οποίους έρχεται σε επικοινωνία, είναι έγκυρες. Σε κάθε διεύθυνση που λαμβάνει ο i, με βάση την Εξ. 3.3 θα περιέχεται το K_i. Αν κάποιος κακόβουλος κόμβος χρησιμοποιήσει κάποια άλλη διεύθυνση με σκοπό να ξεγελάσει τον i, τότε ο i μπορεί εύκολα να ελέγξει την ύπαρξη του K_i στη διεύθυνση που λαμβάνει, εκτελώντας λογικό AND μεταξύ του $BF(K_i)$ και του φίλτρου που έλαβε. Αν το K_i υπάρχει τότε το αποτέλεσμα του AND θα είναι το ίδιο το $BF(K_i)$, άρα η διεύθυνση που έλαβε είναι έγκυρη.

---

[6] Ο τρόπος που προφυλάσσεται η ανωνυμία με αυτό τον τρόπο περιγράφεται στην ενότητα 3.4

Για να γίνουν πιο σαφή τα παραπάνω θεωρούμε το παρακάτω παράδειγμα. Έστω το δίκτυο του Σχήματος 3.1. Συνολικά, έχουμε N = 5 κόμβους και N* = 2 εικονικούς κόμβους, τον F και τον G. Αυτό σημαίνει ότι θα δημιουργηθούν αρχικά $N^2 = 49$ κλειδιά από την ΤΑ. Τα κλειδιά αυτά θα είναι τα:

{K_{AB}, K_{AC}, K_{AD}, K_{AE}, K_{AF}, K_{AG}, K_A, K_{BA}, K_{BC}, K_{BD}, K_{BE}, K_{BF}, K_{BG}, K_B,
K_{CA}, K_{CB}, K_{CD}, K_{CE}, K_{CF}, K_{CG}, K_C, K_{DA}, K_{DB}, K_{DC}, K_{DE}, K_{DF}, K_{DG}, K_D,
K_{EA}, K_{EB}, K_{EC}, K_{ED}, K_{EF}, K_{EG}, K_E, K_{FA}, K_{FB}, K_{FC}, K_{FD}, K_{FE}, K_{FG}, K_F,
K_{GA}, K_{GB}, K_{GC}, K_{GD}, K_{GE}, K_{GF}, K_G }



Σχήμα 3.1 Περιγραφή της πληροφορίας που λαμβάνει ένας κόμβος από την ΤΑ

Με βάση τα παραπάνω κλειδιά η ΤΑ θα δημιουργήσει και θα μοιράσει στους κόμβους τις νέες διευθύνσεις τους. Στο Σχήμα 3.1 βλέπουμε τις διευθύνσεις που θα δημιουργηθούν για τον κόμβο A. Οι 6 πρώτες αφορούν στον τρόπο που ο A θα αναπαριστά τον εαυτό του όταν έρχεται σε επαφή με καθέναν από τους υπόλοιπους κόμβος. Έτσι, π.χ. όταν ο A στέλνει στον B τη λίστα επαφών του, η διεύθυνση αποστολέα που θα μπει στο πακέτο θα είναι η id_A^B. Στο σχήμα φαίνονται ποια από τα κλειδιά δεν περιλαμβάνονται στα id (εμφανίζονται σαν διαγραμμένα). Οι επόμενες 6 σχετίζονται με την αποστολή πακέτων από τον A προς τους υπόλοιπους κόμβους. Συνεπώς, όταν ο A θέλει να στείλει π.χ. στον D ένα πακέτο δεδομένων, τότε η διεύθυνση προορισμού του πακέτου θα είναι η $dest_A^D$. Τέλος, οι διευθύνσεις της

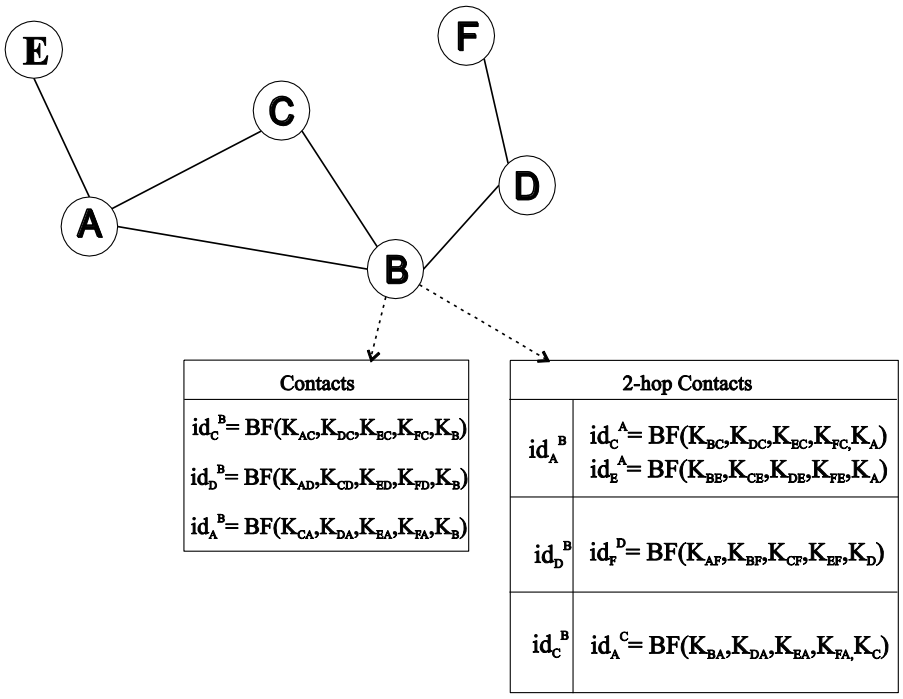
μορφής src_i^j χρησιμεύουν στο να αναγνωρίσει ένας κόμβος τον αποστολέα των πακέτων που λαμβάνει. Όταν ο D στέλνει ένα πακέτο στον A, η διεύθυνση που θα περιέχεται στο πακέτο θα είναι το $BF(K_{DA})$. Εκτελώντας λογικό AND μεταξύ $BF(K_{DA})$ και του src_A^D ο A θα αναγνωρίσει ότι ο αποστολέας είναι ο D. Παρόμοια, οι κόμβοι B, C, D, E θα λάβουν τις αντίστοιχες διευθύνσεις τους.

Αυτή η αναπαράσταση αναμένουμε να επηρεάσει ως ένα βαθμό τον υπολογισμό των μετρικών betweenness και similarity, λόγω της ύπαρξης false positives, όπως έχουμε ήδη αναφέρει. Η επίδραση των false positives στη δρομολόγηση αναλύεται στην ενότητα 3.5. Αντίθετα, η ύπαρξη εικονικών κλειδιών δεν επηρεάζει τους υπολογισμούς σε καμία περίπτωση, αφού τα έγκυρα κλειδιά δεν αφαιρούνται από τα πραγματικά id (δεν λείπει χρήσιμη πληροφορία) και η επικοινωνία θα γίνεται μόνο μεταξύ πραγματικών κόμβων (δεν πρόκειται να έχουμε πακέτα από και προς εικονικούς κόμβους). Το μόνο που χρειάζεται είναι προσεκτικός ορισμός των συνθηκών που πρέπει να ισχύουν κατά την εκτέλεση των δυαδικών λογικών πράξεων μεταξύ των φίλτρων, με τις οποίες θα πραγματοποιούνται οι υπολογισμοί των μετρικών betweenness και similarity. Στην επόμενη ενότητα περιγράφουμε πως πραγματοποιούμε τον υπολογισμό των μετρικών με χρήση των φίλτρων Bloom.

3.3.2. Υπολογισμός betweenness

Για τον υπολογισμό του betweenness, στον παραδοσιακό αλγόριθμο, ένας κόμβος A θα εξέταζε αν υπάρχουν επαφές του, που να μην υπάρχουν στη λίστα επαφών των γειτόνων του. Η ύπαρξη τέτοιων επαφών σημαίνει ότι ο A βρίσκεται ως ενδιάμεσος κόμβος σε συντομότερα μονοπάτια μεταξύ των γειτόνων του, άρα έχει μεγαλύτερη τιμή betweenness. Στον SimBet-BF ο έλεγχος αυτός πραγματοποιείται με λογικές πράξεις (λογικό AND) μεταξύ φίλτρων Bloom. Συγκεκριμένα, για κάθε φίλτρο $Contact_i = BF_i(K_{j,i},...,K_A), \forall j \neq i, A$, που αναπαριστά μια επαφή του A, εφαρμόζεται AND με κάθε φίλτρο $Contacts_{BF}(i)$ που περιέχει τις επαφές του γείτονα i. Αν το φίλτρο που προκύπτει από το λογικό AND έχει $k\times(N-x)$ ή περισσότερα bits ίσα με 1 σημαίνει ότι η επαφή αυτή, με μεγάλη πιθανότητα, υπάρχει και στη λίστα επαφών του i. Με x συμβολίζουμε το πλήθος των κλειδιών που αφαιρούνται κατά την κατασκευή των φίλτρων (τα εικονικά κλειδιά και το K_{vi} που περιγράψαμε στην προηγούμενη

παράγραφο) και με k το πλήθος των συναρτήσεων κατακερματισμού του φίλτρου. Αν η συνθήκη αυτή δεν ισχύει, με βεβαιότητα μπορούμε να πούμε ότι η επαφή αυτή δεν περιέχεται, οπότε ο A ενημερώνει την τιμή του betweenness. Η παραπάνω συνθήκη προκύπτει από το πλήθος των κλειδιών που παραμένουν ίδια στις διευθύνσεις με τις οποίες ο κάθε κόμβος γίνεται γνωστός στους υπόλοιπους (βλέπε Σχήμα 3.1). Κάθε κλειδί που κατακερματίζεται δίνει k θέσεις άσσων. Κάθε κόμβος που έχει συναντήσει τον i, έχει στη λίστα των επαφών του ένα φίλτρο Bloom που τον αναπαριστά και το οποίο θα περιέχει N-x κοινά κλειδιά με οποιοδήποτε από τα υπόλοιπα id_i^j. Συνεπώς, τα κοινά bits που θα είναι άσσοι σε δύο διαφορετικές αναπαραστάσεις του ίδιου κόμβου i θα είναι k×(N-x).



Σχήμα 3.2 Υπολογισμός betweenness

Στο Σχήμα 3.2 βλέπουμε τη διαδικασία που περιγράψαμε με ένα παράδειγμα. Για λόγους απλότητας στο σχήμα και στις δομές δεν περιλαμβάνονται οι εικονικοί κόμβοι και τα εικονικά κλειδιά που παράγει η TA. Άλλωστε ο υπολογισμός του betweenness δεν επηρεάζεται από την ύπαρξή τους, αφού ποτέ δεν πρόκειται μη έγκυροι κόμβοι να εμφανιστούν ως επαφές κάποιων έγκυρων κόμβων.

Ο κόμβος Β έχει αποθηκευμένες τις επαφές του σε μια λίστα (Contacts). Σε μια ξεχωριστή λίστα αποθηκεύει τις επαφές που του έχουν στείλει οι γείτονές του (2-hop Contacts). Κάθε γείτονας στέλνει στον Β τις επαφές του με τη μορφή ενός φίλτρου Bloom που είναι η ένωση των επιμέρους φίλτρων διευθύνσεων των κόμβων που έχει συναντήσει. Για παράδειγμα ο Α θα στείλει στον Β τις επαφές του με τη μορφή:

$$Contacts(A) = \cup id_x^A = \cup BF(\sum_{z \neq A} K_{zx} + K_A)$$

Για να υπολογίσει το betweenness θα πρέπει να ελέγξει ποιες από τις επαφές του υπάρχουν στη λίστα επαφών των γειτόνων του. Έτσι, εφαρμόζει λογικό AND για καθεμία από τις επαφές του με καθεμιά από τις γειτονιές των επαφών του. Στο παράδειγμά μας ο Β αρχικά εκτελεί id_C^B (AND) Contacts(A). Έτσι ελέγχει αν ο id_C^B είναι γείτονας του id_A^B. Ο έλεγχος αυτός θα δώσει θετική απάντηση, μιας και ο αριθμός των κοινών κλειδιών που υπάρχουν στα δύο φίλτρα ικανοποιεί τη συνθήκη που διατυπώσαμε πριν (k·(N-x) κοινά κλειδιά). Το αποτέλεσμα είναι σωστό, αφού όντως ο C είναι κοινός γείτονας του Α και του Β. Στη συνέχεια, θα εκτελέσει id_C^B (AND) id_F^D. Εδώ, ο έλεγχος θα δώσει αρνητική απάντηση, επειδή δεν υπάρχουν κοινά κλειδιά στα δύο φίλτρα. Και σε αυτή την περίπτωση η απάντηση είναι σωστή, μιας και ο C δεν είναι κοινός γείτονας του Β και του D. Αυτό σημαίνει ότι ο Β βρίσκεται σε συντομότερο μονοπάτι μεταξύ C και D, άρα πρέπει να ενημερώσει την τιμή του betweenness.

Αντίστοιχος έλεγχος γίνεται και για τις υπόλοιπες επαφές του Β, ενημερώνοντας όταν πρέπει την τιμή του betweenness.

Ο υπολογισμός των μετρικών, όπως έχουμε αναφέρει, μπορεί να επηρεαστεί από την ύπαρξη false positives. Η πιθανότητα να λάβουμε λανθασμένα θετική απάντηση κατά τον έλεγχο ύπαρξης ενός στοιχείου αποδεικνύεται ότι είναι

$$P_{fp} = (1 - (1 - \frac{1}{m})^{k \cdot n})^k$$

όπου m το μέγεθος του φίλτρου, k ο αριθμός των συναρτήσεων κατακερματισμού και n το πλήθος των στοιχείων που περιέχονται στο φίλτρο. Στην περίπτωσή μας το φίλτρο κατασκευάζεται ώστε να υποστηρίζει N^2 στοιχεία, δηλ. όσα είναι τα κλειδιά

που μπορεί να λάβει ένας κόμβος μέσω των επαφών του. Άρα, η παραπάνω πιθανότητα γράφεται

$$P_{fp} = (1-(1-\frac{1}{m})^{k\cdot N^2})^k$$

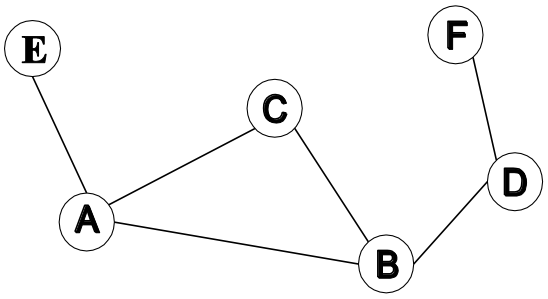
Αυτό σημαίνει ότι όταν εξετάζουμε την ύπαρξη ενός κλειδιού στο φίλτρο, με πιθανότητα p_{fp} θα λάβουμε μια λανθασμένα θετική απάντηση. Κατά τον υπολογισμό του betweenness, εξετάζουμε αν μια επαφή (που αναπαρίσταται με ένα φίλτρο Bloom) υπάρχει στη λίστα επαφών ενός κόμβου. Η συνθήκη που θέτουμε για την ύπαρξη ενός κόμβου σε ένα φίλτρο είναι να υπάρχουν N-x κοινά κλειδιά. Αυτό ο έλεγχος είναι ισοδύναμος με το membership check για την ύπαρξη ενός κόμβου σε ένα φίλτρο επαφών. Αυτό σημαίνει ότι ένας έλεγχος ύπαρξης ενός κόμβου θα σώσει ή κανένα κοινό κλειδί ή ένα πλήθος κλειδιών μεγαλύτερο ή ίσο από N-x. Η πιθανότητα να λάβουμε μια λανθασμένα θετική απάντηση είναι να μην υπάρχει κανένα κοινό κλειδί και λόγω false positives να λάβουμε N-x κοινά κλειδιά, Συνεπώς, η πιθανότητα να έχουμε λάθος απάντηση κατά τον υπολογισμό του betweenness είναι

$$P^*_{fp} = (1-(1-\frac{1}{m})^{k\cdot N^2})^{k\cdot(N-x)}$$

δηλ. $(P_{fp})^{k\cdot(N-x)}$ η οποία είναι πολύ μικρότερη από την P_{fp}. Η πιθανότητα αυτή σημαίνει ότι στην πράξη δεν πρόκειται να έχουμε λάθος υπολογισμό betweenness.

3.3.3. Υπολογισμός similarity

Αντίστοιχος υπολογισμός γίνεται και για την μετρική similarity (Σχήμα 3.3). Για κάθε προορισμό dest για τον οποίο ένας κόμβος B έχει αποθηκευμένα πακέτα εκτελεί λογικό AND μεταξύ του κάθε φίλτρου επαφών που έχει λάβει από τις επαφές του (2-hop contacts) και του φίλτρου που αναπαριστά τον προορισμό ($dest_i^j$). Σε αυτή την περίπτωση το πλήθος των bits που πρέπει να έχουν τεθεί 1 ισούται με k, όσο δηλ. και το πλήθος των συναρτήσεων κατακερματισμού. Αυτό συμβαίνει επειδή το φίλτρο Bloom που αναπαριστά τον προορισμό σε ένα πακέτο περιέχει μόνο ένα κλειδί (βλέπε Σχήμα 3.1). Αν η παραπάνω συνθήκη ισχύει, σημαίνει ότι κόμβος στου οποίου το φίλτρο επαφών περιέχεται ο προορισμός, είναι κοινή επαφή του B και του προορισμού. Σε αυτή την περίπτωση ο B πρέπει να ενημερώσει την τιμή του similarity για τον προορισμό d.



Destinations
$dest_{FE} = BF(K_{FE})$
$dest_{BA} = BF(K_{BA})$

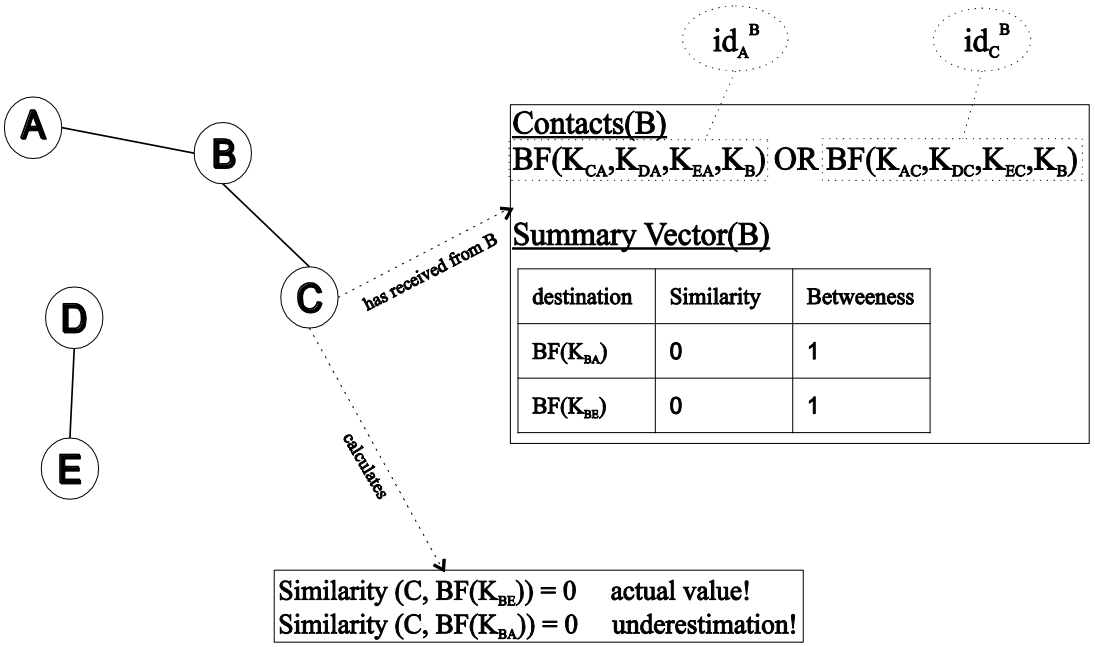
2-hop Contacts	
id_A^B	$id_C^A = BF(K_{BC},K_{DC},K_{EC},K_{FC},K_A)$ $id_E^A = BF(K_{BE},K_{CE},K_{DE},K_{FE},K_A)$
id_D^B	$id_F^D = BF(K_{AF},K_{BF},K_{CF},K_{EF},K_D)$
id_C^B	$id_A^C = BF(K_{BA},K_{DA},K_{EA},K_{FA},K_C)$

Σχήμα 3.3 Υπολογισμός similarity

Στο Σχήμα 3.3 βλέπουμε τον υπολογισμό του similarity με ένα παράδειγμα. Για λόγους απλότητας και πάλι δεν συμπεριλαμβάνουμε τους εικονικούς κόμβους και τα εικονικά κλειδιά στο σχήμα. Παρατηρούμε, λοιπόν, ότι το Β διαθέτει πακέτα για δύο προορισμούς, τον Ε ($dest_{FE}$) και τον Α ($dest_{BA}$). Στο παράδειγμά μας, λοιπόν, ο Β θα εκτελέσει λογικό AND αρχικά με το $dest_{FE}$ και καθένα από τα φίλτρα επαφών των γειτόνων του. Μόνο το φίλτρα επαφών του id_A^B θα δώσει θετική απάντηση, που σημαίνει ότι ο Β έχει similarity = 1 με τον προορισμό του πακέτου. Όπως, βλέπουμε ο μπορεί να κάνει αυτόν τον υπολογισμό χωρίς να ξέρει ποιος είναι ο πραγματικός προορισμός του πακέτου ούτε ποιος κόμβος είναι η κοινή επαφή τους. Για τον $dest_{BA}$ ο έλεγχος θα δώσει θετική απάντηση μόνο για το φίλτρο επαφών του C (id_C^B), που σημαίνει ότι ο Β έχει ως κοινή επαφή με τον προορισμό τον id_C^B.

Κατά τον υπολογισμό του similarity φαίνεται καθαρά ο λόγος ύπαρξης δύο κλειδιών ανά ζεύγος επικοινωνίας. Αν χρησιμοποιούσαμε το ίδιο κλειδί για καθεμία από τις κατευθύνσεις επικοινωνίας, ο υπολογισμός του similarity θα ήταν προβληματικός. Για τον ίδιο παραλήπτη (τον Ε), ο Β θα έπαιρνε θετική απάντηση για την ύπαρξη του $dest_{FE}$, τόσο κατά τον έλεγχο με το id_E^A, όσο και με το id_F^D, επειδή δοθέντος ενός και

μόνο συμμετρικού κλειδιού δεν θα μπορούσε να ξεχωρίζει ποιος είναι ο πραγματικός προορισμός. Γενικεύοντας, μπορούμε να πούμε ότι με την ύπαρξη ενός κλειδιού ανά ζεύγος δεν μπορούμε να «ξεχωρίσουμε» ποιος κόμβος από το ζεύγος είναι ο αποστολέας και ποιος ο παραλήπτης.



Σχήμα 3.4 Περίπτωση υποεκτίμησης της μετρικής similarity

Ένα άλλο ζήτημα με τον υπολογισμό του similarity είναι η πιθανότητα να έχουμε υποεκτίμηση της πραγματικής τιμής της, η οποία οφείλεται στον τρόπο που έχουμε κατασκευάσει τα φίλτρα. Συγκεκριμένα, υπενθυμίζουμε ότι για λόγους προστασίας της ανωνυμίας, έχουμε αφαιρέσει το κλειδί K_{ji} από την διεύθυνση με την οποία ο i γίνεται γνωστός στον j. Η αφαίρεση αυτού του κλειδιού μπορεί να οδηγήσει σε υποεκτίμηση της τιμής του similarity.

Η περίπτωση αυτή φαίνεται με το παράδειγμα του Σχήματος 3.4. Στο Σχήμα 3.4 ο Β, έχοντας κάποιο πακέτο προς αποστολή για τον Α, τον περιλαμβάνει ($BF(K_{BA})$) στον summary vector που στέλνει στον C. Ο C έχει λάβει ήδη το φίλτρο επαφών του Β στο οποίο περιλαμβάνεται και το φίλτρο που αναπαριστά τον Α. Ωστόσο, το κλειδί K_{BA} δεν περιλαμβάνεται στο φίλτρο, επειδή έτσι έχει κατασκευαστεί εξ' αρχής το φίλτρο από την ΤΑ. Αυτό έχει ως συνέπεια ο C, εκτελώντας λογικό AND μεταξύ του $BF(K_{BA})$ και του φίλτρου που περιέχει τις επαφές του Β, να μην λαμβάνει θετική

απάντηση σχετικά με την ύπαρξη του στοιχείου στο φίλτρο. Συνεπώς, η τιμή του similarity που υπολογίζει είναι 0 ενώ με τον παραδοσιακό αλγόριθμο θα ήταν 1.

Γενικεύοντας, μπορούμε να πούμε ότι ο υπολογισμός της τιμής του similarity σε έναν κόμβο v για τον κόμβο προορισμό d (Sim(v,d)) οδηγεί σε υποεκτίμηση, αν ο d είναι επαφή του v.

Η παραπάνω περίπτωση είναι αρκετά σπάνια γιατί εμφανίζεται κυρίως όταν κάποιος κόμβος δεν έχει προλάβει να προωθήσει όλα τα πακέτα σε κάποια από τις επαφές του, όσο υπήρχε σύνδεση μεταξύ τους. Το αντίκτυπο στην επίδοση δεν είναι σημαντικό και αυτό φαίνεται στην πειραματική αξιολόγηση της μεθόδου μας στο Κεφάλαιο 4. Μάλιστα, υπάρχουν περιπτώσεις που αυτό το φαινόμενο οδηγεί σε οριακά καλύτερη απόδοση του αλγορίθμου SimBet-BF, όπως θα δούμε στην πειραματική αξιολόγηση του αλγορίθμου. Αυτό οφείλεται στο γεγονός ότι σε αυτές τις περιπτώσεις η καλύτερη στρατηγική είναι το πακέτο να μην προωθηθεί και να παραμείνει στον B (βλέπε Σχήμα 3.4).

3.3.4. Διαδικασία δρομολόγησης

Χρησιμοποιώντας την αναπαράσταση των κόμβων που ορίσαμε την ενότητα 3.3.1, θα περιγράψουμε την τελική μορφή του ανωνυμοποιημένου αλγορίθμου SimBet-BF. Ο ψευδοκώδικας φαίνεται στο Σχήμα 3.5.

Καταρχήν, το πρώτο βήμα του αλγορίθμου δεν μπορεί να παραμείνει ίδιο με τον παραδοσιακό αλγόριθμο. Υπενθυμίζουμε ότι όταν ένας κόμβος A συναντά ένα νέο γείτονα B, αρχικά, ο A παραδίδει στον B όσα από τα αποθηκευμένα μηνύματά του έχουν τον B ως προορισμό. Αυτό το βήμα θα οδηγούσε σε άμεση παραβίαση της ανωνυμίας, αφού ο A θα καταλάβει ότι τελικός παραλήπτης των μηνυμάτων θα είναι ο B. Δεδομένου ότι ο B είναι αυτός που θα αποφασίσει στη συνέχεια για ποιους κόμβους θα του προωθήσει μηνύματα ο A, μπορεί να συμπεριλάβει τον εαυτό του στη λίστα αυτών των κόμβων (request vector). Η υπόλοιπη λειτουργία του αλγορίθμου παραμένει ως έχει σε επίπεδο επικοινωνίας μεταξύ των κόμβων.

Algorithm 1 SimBet-BF Routing Algorithm, pseudo-code of node A

```
1:  upon reception of Hello message h from node B do
2:    if newNeighbour(B) == true
3:      requestEncounters(B)

7:  upon reception of encounter vector ev from node B do
8:    addNodeEncounters(B, ev)
9:    updateBetweenness()
10:   updateSimilarity()
11:   exchangeSummaryVector(B)
12:
13: upon reception of summary vector sv from node B do
14:   Vector requestMsgs
15:   for all destinations ∈ sv do
16:     if B.simBet(d) < simBet(d) or A == d
17:       requestMsgs.add(d)
18:   sendMsgRequest(B, requestMsgs)
19:
20: upon reception of message request vector mrv from node B
    do
21:   Vector transferMsgs
22:   for all messages ∈ mrv do
23:     transferMsgs.add(msgQueue.getMsgs(d))
24:   sendTransferMsgs(B, transferMsgs)
25:
26: upon reception of transfer message tm from node B do
27:   msgQueue.add(tm)
```

Σχήμα 3.5 Ψευδοκώδικας του αλγορίθμου SimBet-BF

Ωστόσο, οι δομές που ανταλλάσσονται στα ενδιάμεσα βήματα του αλγορίθμου δεν είναι οι ίδιες με τις αρχικές.

Ο κόμβος Α, λοιπόν, θα στείλει στον Β κατευθείαν το αίτημα για την αποστολή των επαφών του, χωρίς να παραδώσει πακέτα που προορίζονται γι' αυτόν. Πάνω στο πακέτο της αίτησης θα τοποθετήσει ως διεύθυνση το id_A^B. Με αυτόν τον τρόπο ο Β μπορεί να καταγράψει τον Α ως επαφή του. Ο Β αφού λάβει το αίτημα του Α για αποστολή των επαφών του, δεν θα στείλει την λίστα διευθύνσεων των επαφών του, αλλά μια αναπαράστασή τους με χρήση φίλτρου Bloom. Ο Β θα έχει καταγράψει ως εκείνη τη στιγμή ένα σύνολο επαφών με τη μορφή φίλτρων Bloom. Η μορφή αυτών

των διευθύνσεων είναι: $Contact_i = BF_i(K_{j,i},...,K_B), \forall j \neq i, B$, δηλ. πρόκειται για μια λίστα από φίλτρα Bloom που αναπαριστούν τους κόμβους που έχει συναντήσει. Αντί να στείλει αυτή τη λίστα ως έχει, προτείνουμε να δημιουργήσει ένα νέο φίλτρο Bloom, το οποίο θα προκύψει από την ένωση των επιμέρους φίλτρων της λίστας. Η ένωση είναι μια λειτουργία που μπορεί να πραγματοποιηθεί σε φίλτρα Bloom εφαρμόζοντας λογικό OR (bitwise-OR) μεταξύ τους. Με τον τρόπο αυτό η πληροφορία που περιέχεται σε καθένα από τα επιμέρους φίλτρα συνενώνεται σε ένα, χωρίς να έχουμε απώλεια πληροφορίας.

Κατά τη λειτουργία αυτή πρέπει να προσέξουμε κάποια λεπτά σημεία που μπορεί να οδηγήσουν σε άμεση παραβίαση της ανωνυμίας. Πρώτον, στην λίστα επαφών του Β μπορεί να υπάρχει ο ίδιος ο Α (π.χ. λόγω παλιότερης συνάντησής τους). Σε αυτή την περίπτωση, από την ένωση των φίλτρων-επαφών του Β θα πρέπει να εξαιρεθεί τυχόν υπάρχουσα επαφή που αναφέρεται στον Α. Σε αντίθετη περίπτωση, αν ο Α είναι κακόβουλος, μπορεί να ελέγξει την ύπαρξή του στο νέο φίλτρο επαφών του Β και να καταλάβει ότι ο κόμβος με τον οποίο επικοινωνεί είναι ο Β. Συνεπώς, ο Β πριν εκτελέσει το λογικό OR μεταξύ των φίλτρων-επαφών του πρέπει να ελέγξει ότι ο κόμβος με τον οποίο συνδιαλέγεται δεν περιλαμβάνεται στη λίστα επαφών του. Αυτό μπορεί να γίνει εύκολα, εφαρμόζοντας λογικό XOR μεταξύ του φίλτρου που αναπαριστά τον Α στο πακέτο request encounters και των επαφών του Β. Αν ο Α υπάρχει στη λίστα, το αποτέλεσμα του XOR θα δώσει ένα δυαδικό πίνακα με όλες τις θέσεις του να έχουν τιμή μηδέν, άρα η συγκεκριμένη επαφή θα πρέπει να εξαιρεθεί από την ένωση των φίλτρων. Δεύτερον, όλες οι επαφές του Β θα περιέχουν το ατομικό κλειδί του Β, το K_B. Σε αυτή την περίπτωση, ο Α μπορεί να αναγνωρίσει ότι ο κόμβος με τον οποίο συνδιαλέγεται είναι ο Β, δοκιμάζοντας (με χρήση λογικού AND) με όλα τα id_A^j. Εφαρμόζοντας λογικό AND μεταξύ του φίλτρου επαφών που θα λάβει από τον Β και καθενός από τα id_A^j, μόνο το id_A^B θα ταιριάζει, συνεπώς ο Α θα αναγνωρίσει τον Β. Για το λόγο αυτό, πριν ο Β στείλει το φίλτρο επαφών στον Α εφαρμόζει λογικό XOR μεταξύ του φίλτρου επαφών του και του φίλτρου που αναπαριστά το ατομικό του κλειδί ($BF(K_B)$), το οποίο έχει λάβει από την ΤΑ. Έτσι, τα αντίστοιχα bit στο φίλτρο επαφών μηδενίζονται και αποτρέπεται η άμεση αναγνώρισή του από τον Α.

Το νέο φίλτρο $Contacts_{BF}(B)$, θα έχει το ίδιο μέγεθος με τα φίλτρα από τα οποία προήλθε, όποτε έχουμε και εξοικονόμηση χώρου στην αποστολή του. Επίσης, περιορίζουμε την πιθανότητα, αν ο Α είναι κακόβουλος χρήστης, να συλλέξει πληροφορία από τα επιμέρους φίλτρα και να τη χρησιμοποιήσει για να σπάσει την ανωνυμία. Αναλυτικότερη περιγραφή της πληροφορίας που μπορεί να εξάγει ο Α από τη λίστα επαφών του Β γίνεται στην ενότητα 3.4.

Αφού ο κόμβος Α λάβει το φίλτρο που περιέχει τις επαφές του Β, θα το χρησιμοποιήσει για να υπολογίσει και πιθανόν να ανανεώσει τις τιμές των μετρικών betweenness και similarity, όπως περιγράψαμε στις προηγούμενες παραγράφους.

Αφού ολοκληρωθεί η ενημέρωση των μετρικών, ο κόμβος Β κατασκευάζει τον summary vector (sv), ενσωματώνει τις νέες τιμές των μετρικών για κάθε προορισμό του sv και τον στέλνει στον Α. Όταν ο Α λάβει τον sv, ακολουθεί την ίδια διαδικασία, όπως στο παραδοσιακό αλγόριθμο, συγκρίνοντας την τιμή του SimBetUtil του Β με το δικό του, για κάθε προορισμό του sv. Οι προορισμοί για τους οποίους ο Α έχει μεγαλύτερο SimBetUtil εισάγονται σε μια λίστα message request vector (mrv). Όπως αναφέραμε παραπάνω, εκτός από αυτούς τους προορισμούς, αν στον sv περιέχεται ο ίδιος ο Α, ανεξαρτήτως τιμής του SimBetUtil, ο Α θα εισάγει τον εαυτό του στη λίστα mrv. Αυτός είναι ο μόνος τρόπος να παραλάβει τα πακέτα που δεν παραδόθηκαν άμεσα κατά την έναρξη της επικοινωνίας μεταξύ των δύο κόμβων.

Τέλος, ο Α στέλνει τη λίστα mrv με τους προορισμούς στον Β. Αυτός προωθεί όσα από τα πακέτα που διαθέτει έχουν ως προορισμό κάποιον από τους προορισμούς που περιέχονται στον mrv.

Η δρομολόγηση με τον παραπάνω τρόπο μας εξασφαλίζει ότι οι κόμβοι μπορούν να επικοινωνούν ανώνυμα. Η αναπαράσταση με χρήση φίλτρων Bloom βοηθά στο να αποκρύψουμε την πληροφορία των επαφών από τους ενδιάμεσους κόμβους καθώς και τον αποστολέα/παραλήπτη ενός πακέτου. Κανένας ενδιάμεσος κόμβος δεν μπορεί να σπάσει άμεσα την ανωνυμία, καθώς δεν μπορεί να εξάγει κάποια πληροφορία από

το περιεχόμενο του πακέτου, ούτε από την πληροφορία που λαμβάνει από τον εκάστοτε νέο γείτονά του. Επίσης, ο υπολογισμός των μετρικών επιτυγχάνεται αποτελεσματικά, με απλές λογικές πράξεις μεταξύ των φίλτρων (δυαδικών πινάκων). Επιπλέον, τη δημιουργία των φίλτρων-διευθύνσεων αναλαμβάνει μια αξιόπιστη οντότητα και έτσι οι κόμβοι δεν έχουν καμία γνώση σχετικά με τα αρχικά κλειδιά που κατακερματίστηκαν για τη κατασκευή τους. Στην επόμενη ενότητα εξετάζουμε τις αντοχές του μοντέλου σε επιθέσεις από κακόβουλους χρήστες.

3.4. Εξασφάλιση ανωνυμίας και αντοχή του μοντέλου σε επιθέσεις

Σε αυτή την υποενότητα θα εξετάσουμε τις πιθανές απειλές κατά της ανωνυμίας που μπορεί να παρουσιαστούν στο μοντέλο μας και πως αυτές αντιμετωπίζονται. Οι βασικοί στόχοι που πρέπει να πληρούνται είναι η προστασία της ανωνυμίας κατά την άμεση επικοινωνία δύο κόμβων και η διατήρηση της ανωνυμίας μέσω της ανταλλαγής των επαφών.

3.4.1. Εξασφάλιση ανωνυμίας κατά την άμεση επαφή

Όπως αναφέραμε στην παράγραφο 3.3.1, η άμεση συσχέτιση του id ενός κόμβου, έτσι όπως αυτό στέλνεται σε κάποιο γείτονά του, με την πραγματική του διεύθυνση είναι αδύνατη. Υπενθυμίζουμε ότι από κάθε φίλτρο της μορφής id_i^v, το οποίο αναφέρεται στον τρόπο με τον οποίο ο κόμβος i γίνεται γνωστός στον v, δεν περιλαμβάνεται το κλειδί K_{vi}. Σε αυτή την περίπτωση ο κόμβος v δεν μπορεί να χρησιμοποιήσει με κανένα τρόπο την πληροφορία που του παρέχει η ΤΑ (συγκεκριμένα τα φίλτρα $dest_v^j$) για να αποκαλύψει ποια είναι η πραγματική διεύθυνση του κόμβου που του στέλνει ως διεύθυνση του id_i^v.

Στη διατήρηση της ανωνυμίας κατά την άμεση επικοινωνία δύο κόμβων συμβάλλει και η ύπαρξη εικονικών κόμβων και κλειδιών. Υπενθυμίζουμε ότι η ΤΑ παράγει ids που δεν σχετίζονται με πραγματικούς κόμβους του δικτύου τα οποία όμως μοιράζονται κανονικά στους κόμβους, χωρίς αυτοί να μπορούν να διακρίνουν τα πραγματικά από τα εικονικά. Με βάση το Σχήμα 3.1 μπορούμε να εξηγήσουμε πως μας χρησιμεύουν τα επιπλέον εικονικά κλειδιά και διευθύνσεις που παράγονται από

την ΤΑ. Ας πάρουμε για παράδειγμα τον κόμβο Α και ας υποθέσουμε ότι είναι κακόβουλος. Γνωρίζοντας ότι το src_A^B αναφέρεται στον κόμβο Β, μπορεί να ελέγξει ποιο από τα id_A^j που διαθέτει περιέχει το K_{BA}. Αν τα εικονικά id_A^j δεν υπήρχαν, τότε εύκολα θα μπορούσε να αποκαλύψει το πραγματικό id με το οποίο σχετίζεται, επειδή, λόγω της κατασκευής των φίλτρων των διευθύνσεων το κλειδί που περιέχεται σε οποιοδήποτε src_A^j περιέχεται σε όλα τα id εκτός από id_A^j. Στην περίπτωση του παραδείγματος, προσθέτοντας εικονικούς κόμβους και αφαιρώντας από τα id τους έγκυρα κλειδιά, σημαίνει ότι εκτός από το id_A^B, υπάρχουν και κάποια από τα εικονικά id που δεν θα περιέχουν το src_A^B (στην περίπτωσή μας το id_A^F). Άρα ο Α μπορεί να υποθέσει ότι είτε το id_A^B είτε το id_A^F αφορά τον κόμβο Β. Αν δεν χρησιμοποιούσαμε εικονικά id η πιθανότητα αυτή θα ήταν 100%, αφού μόνο το id_A^B δεν θα περιείχε το K_{BA}. Συνεπώς, αν δεν υπήρχε το id_A^F ο Α θα μπορούσε να είναι σίγουρος ότι το id_A^B είναι η διεύθυνση του Β, σπάζοντας την ανώνυμη επικοινωνία. Μάλιστα, στην περίπτωση που το πλήθος των εικονικών κόμβων είναι ίσο με το πλήθος των πραγματικών κόμβων στο δίκτυο, τότε η πιθανότητα ο Α να προβλέψει σωστά την πραγματική ταυτότητα του κόμβου με τον οποίο συνδιαλέγεται είναι 1/(N+1), που είναι το καλύτερο που μπορούμε να έχουμε.

3.4.2. Εξασφάλιση ανωνυμίας κατά την ανταλλαγή πακέτων

Ένα άλλο θεμελιώδες ζήτημα ανωνυμίας σε DTN/opportunistic δίκτυα, είναι η προστασία της ανωνυμίας στο τελευταίο άλμα της επικοινωνίας. Ειδικότερα, ο κόμβος που θα παραδώσει ένα πακέτο στον τελικό παραλήπτη αναπόφευκτα θα μάθει την ταυτότητά του, αποκαλύπτοντας το ένα από τα δύο άκρα της επικοινωνίας. Το ζήτημα αυτό το αντιμετωπίζουμε επιτυχώς αλλάζοντας το πρώτο βήμα του αλγόριθμου και καταργώντας την άμεση παράδοση των πακέτων με βάση την πραγματική διεύθυνση του παραλήπτη. Ο κόμβος που θα παραδώσει το πακέτο στον τελικό παραλήπτη θα μάθει μόνο την αναπαράσταση (φίλτρο Bloom) της διεύθυνσης του προορισμού και όχι την πραγματική του ταυτότητα.

Ένα άλλο ζήτημα που αφορά την ανταλλαγή πληροφορίας είναι η άμεση αποκάλυψη είτε του αποστολέα είτε του παραλήπτη ενός πακέτου σε οποιονδήποτε ενδιάμεσο κόμβο που θα αναλάβει την προώθησή του. Η αναπαράσταση του προορισμού στο

πακέτο με ένα φίλτρο Bloom, κάνει αδύνατη την αυτή την αποκάλυψη. Μόνο ο τελικός παραλήπτης είναι αυτός που θα μάθει ότι το πακέτο προορίζεται γι' αυτόν. Η μόνη πληροφορία που μπορεί να πάρει ένας ενδιάμεσος κόμβος προέρχεται από τον υπολογισμό του similarity και αυτό που μαθαίνει είναι ότι ο προορισμός βρίσκεται στη 2-hop γειτονιά του, χωρίς ωστόσο να ξέρει ούτε ποιος είναι ο τελικός προορισμός, ούτε μέσω ποιας λίστας επαφών (ποιου κόμβου η λίστα) σχετίζεται με τον προορισμό.

3.4.3. Εξασφάλιση ανωνυμίας κατά την ανταλλαγή των επαφών

Ένα άλλο σημαντικό ζήτημα είναι το αν μπορεί ένας κακόβουλος χρήστης να εξάγει κάποια γνώση συνδυάζοντας πληροφορία από τις επαφές που έχει, τα φίλτρα επαφών των κόμβων που συναντά και από τα πακέτα που προορίζονται γι' αυτόν ή που αναλαμβάνει να προωθήσει. Για να εξετάσουμε αυτή την περίπτωση πρέπει να δούμε τι είδους πληροφορία μπορεί να συλλέξει ένας κόμβος και πως μπορεί να συνδυαστεί με την πληροφορία που λαμβάνει αρχικά από την ΤΑ.

Στον Πίνακα 3.1 φαίνεται τι πληροφορία διαθέτει εξ' αρχής ένας κόμβος Α (από την ΤΑ) και τι μπορεί να μάθει κατά την ανταλλαγή πληροφορίας με τους άλλους κόμβους. Ουσιαστικά μπορούμε να δούμε ότι η πληροφορία που γνωρίζει είναι ό,τι του παρέχει η ΤΑ κατά την έναρξη λειτουργίας του δικτύου, όπως αυτά παρουσιάζονται στην ενότητα 3.3.1.

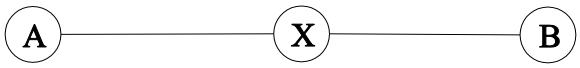
Πίνακας 3.1 Πληροφορία που γνωρίζει/μαθαίνει ο Α

Γνωρίζει	**Μαθαίνει**
$id_A^x = BF(\sum_{i \neq x,A} K_{iA} + K_x)$	$id_y^A = BF(\sum_{i \neq y,A} K_{iy} + K_A)$
$dest_A^x = BF(K_{Ax})$	$Contacts_y^A = \cup_{j \in Contacts(y)} BF(\sum_{i \neq j,y} K_{ij} + K_y)$
$src_A^x = BF(K_{xA})$	
$BF(K_A)$	

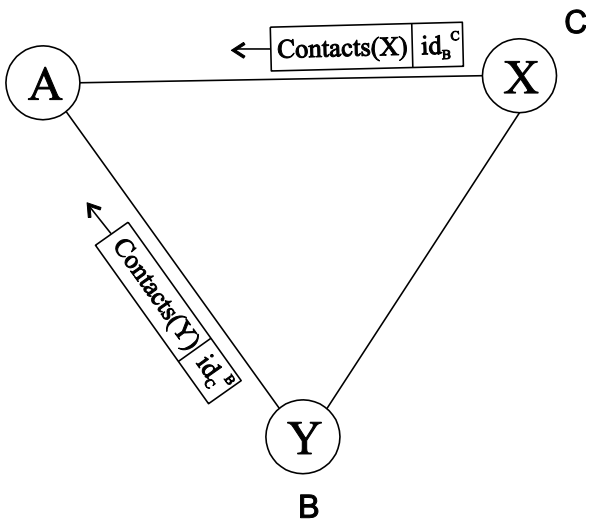
Η επιπλέον πληροφορία που μαθαίνει αφορά τις διευθύνσεις που θα περιέχονται πάνω στα πακέτα που θα ανταλλάσει με τους κόμβους που συναντά, κατά την εκτέλεση του πρωτοκόλλου (request encounters, summary vector κ.λπ.). Επίσης, μπορεί να μάθει τις λίστες επαφών των γειτόνων του (πληροφορία απαραίτητη για σωστή λειτουργία της δρομολόγησης), όχι όμως μεμονωμένα, αλλά ως ένωση (λογικό OR) των επιμέρους επαφών κάθε γείτονα.

Με βάση τον Πίνακα 3.1 μπορούμε να δούμε ότι κάποιος άμεσος συνδυασμός πληροφορίας που να παραβιάζει την ανωνυμία δεν μπορεί να υπάρξει. Ο κόμβος A δεν μπορεί να μάθει άμεσα ποια είναι πραγματική διεύθυνση του κόμβου με τον οποίο συνδιαλέγεται, αφού τα id_y^A δεν μπορούν να αποκαλύψουν πληροφορία σχετικά με τον y. Κανένας συνδυασμός είτε με τα src_A^x, είτε με τα id_A^x, είτε με τα $dest_A^x$ δεν δίνει ταίριασμα με τις διευθύνσεις id_y^A που λαμβάνει ο A από τις επαφές του. Άρα άμεση παραβίαση της ανωνυμίας δεν μπορεί να υπάρξει.

Αντίθετα, παρατηρώντας προσεκτικά τον Πίνακα 3.1 συμπεραίνουμε ότι ο κόμβος A μπορεί να εξάγει πληροφορία από τις επαφές που λαμβάνει σχετικά με την 2-hop γειτονιά των επαφών του. Χρησιμοποιώντας τα $dest_A^j$ και ελέγχοντας με τα φίλτρα επαφών που μαθαίνει από τους κόμβους με τους οποίους επικοινωνεί (χρήση λογικού AND) μπορεί να μάθει ποιοι κόμβοι βρίσκονται στη 2-hop γειτονιά του. Εξάγει δηλ. πληροφορία της μορφής:



Σε αυτή την περίπτωση το λογικό AND μεταξύ $dest_A^C$ και του φίλτρου επαφών του C θα δώσει το ίδιο το $dest_A^C$ που σημαίνει ότι ο C υπάρχει στη λίστα επαφών του X. Άρα, ο A μαθαίνει ότι μέσω ενός άγνωστου κόμβου ο οποίος του στέλνει τις επαφές του, έχει στη 2-hop γειτονιά του τον C. Αυτές οι συσχετίσεις είναι απαραίτητες για τη σωστή λειτουργία υπολογισμού του similarity. Τέτοιου είδους γνώση, όμως, μπορεί να είναι χρήσιμη για έναν κακόβουλο χρήστη σε περιπτώσεις όπου οι κόμβοι που εμπλέκονται σχηματίζουν πλήρη τοπολογία.



Σχήμα 3.6 Πλήρης τοπολογία με τρεις κόμβους

Στο Σχήμα 3.6 βλέπουμε ένα τέτοιο παράδειγμα. Ο Α θα λάβει το φίλτρο επαφών από τους Χ και Υ (που αντιστοιχούν στους C και B), των οποίων τα πραγματικά id είναι άγνωστα. Ωστόσο, με βάση τις επαφές που θα λάβει, μπορεί να καταλάβει ότι μέσω του Υ έχει 2-hop επαφή τον C και μέσω του Χ έχει 2-hop επαφή τον B. Σε αυτή την περίπτωση εάν ο Α γνώριζε το πλήθος των επαφών που περιέχονται σε κάθε ένα από τα φίλτρα που λαμβάνει θα μπορούσε με απόλυτη σιγουριά να εξάγει το συμπέρασμα ότι ο Χ είναι ο C και ο Υ είναι ο B. Αυτήν την πληροφορία ο Α δεν μπορεί να την έχει. Υπενθυμίζουμε ότι στο φίλτρο επαφών που στέλνει κάθε κόμβος δεν περιέχονται μεμονωμένα φίλτρα, αλλά η ένωσή τους. Επίσης, η ΤΑ έχει φροντίσει να αφαιρέσει κάποια από τα μη έγκυρα κλειδιά στα id που μοιράζει αρχικά στους κόμβους. Ο Α, γνωρίζοντας ότι κάποια κλειδιά λείπουν (χωρίς όμως να ξέρει ποια), δεν μπορεί να είναι σίγουρος για το πλήθος των επαφών που περιέχονται στο φίλτρο που λαμβάνει. Για παράδειγμα, εκτός από τον κόμβο C που «βλέπει» μέσω του Υ θα μπορούσε να υπάρχει και μια άλλη επαφή z στο φίλτρο του Υ, από την οποία έχει αφαιρεθεί το κλειδί K_{AZ} και συνεπώς ο Α δεν μπορεί να την εντοπίσει με το $dest_A^Z = BF(K_{AZ})$. Βέβαια, αυτό ισχύει μόνο για τους εικονικούς κόμβους, όμως ο Α δεν είναι σε θέση να διαχωρίσει ποιοι είναι εικονικοί και ποιοι όχι. Η παραπάνω περίπτωση είναι η πιο απλή σε πλήρη τοπολογία. Ωστόσο, το σκεπτικό γύρω από τον τρόπο με τον οποίο προστατεύεται τελικά η ταυτότητα των κόμβων είναι το ίδιο και σε πλήρεις τοπολογίες με περισσότερους κόμβους.

Συνεπώς, βλέπουμε ότι η πληροφορία που περιέχεται στα φίλτρα έχει επιλεγεί με τέτοιο τρόπο, ώστε οι κακόβουλοι κόμβος να μη μπορούν να σπάσουν την ανωνυμία είτε άμεσα είτε έμμεσα. Κανένας συνδυασμός της πληροφορίας που απεικονίζεται στον Πίνακα 3.1 δεν μπορεί να συσχετίσει τα id_i^j με τις πραγματικές διευθύνσεις των κόμβων. Αντίστοιχα, κατά την προώθηση των πακέτων δεν μπορεί να αποφασιστεί με βεβαιότητα ποιος είναι ο αποστολέας και ποιος ο παραλήπτης ενός πακέτου. Η πληροφορία που περιέχεται στο πακέτο είναι η ελάχιστη δυνατή, ώστε να εξασφαλίζεται τόσο ο σωστός υπολογισμός των μετρικών (χωρίς να έχουμε γνώση των πραγματικών διευθύνσεων), όσο και η ανώνυμη από άκρο σε άκρο επικοινωνία.

3.5. Σχέση αποδοτικότητας δρομολόγησης/ιδιωτικότητας

Η χρήση φίλτρων Bloom, λοιπόν, αποτελεί μια εναλλακτική λύση για εξασφάλιση ιδιωτικότητας κατά τη δικτυακή επικοινωνία σε ένα opportunistic δίκτυο, αποφεύγοντας την κρυπτογράφηση. Ωστόσο, τα φίλτρα Bloom, λόγω της κατασκευής τους, δημιουργούν ένα ζήτημα σχετικά με την αποδοτικότητα της δρομολόγησης, το οποίο χρήζει μελέτης. Τα φίλτρα Bloom υπάρχει πιθανότητα να δώσουν λάθος απάντηση στη λειτουργία ελέγχου ύπαρξης ενός στοιχείου (membership check). Συγκεκριμένα, υπάρχει πιθανότητα να λάβουμε μια λανθασμένα θετική απόκριση (false positive), δηλ. ότι μια επαφή ή ένας προορισμός υπάρχει σε ένα φίλτρο, ωστόσο στην πραγματικότητα να μην υπάρχει. Σε αυτή την περίπτωση δημιουργείται ζήτημα στην απόδοση της δρομολόγησης. Τυχόν λάθη που μπορεί να γίνουν οδηγούν σε λάθος υπολογισμούς μετρικών και πιθανόν σε λάθος αποφάσεις δρομολόγησης. Ωστόσο, πιστεύουμε ότι τα λάθη αυτά δεν θα έχουν μεγάλο αντίκτυπο στην επίδοση. Παρακάτω παραθέτουμε μια σειρά παρατηρήσεων που δικαιολογούν την παραπάνω άποψη.

Καταρχάς, ένα φίλτρο Bloom μπορεί να κατασκευαστεί με τέτοιο τρόπο ώστε η πιθανότητα false positive p_{fp} να είναι μικρή (π.χ. 1%). Δοθείσης της πιθανότητας p_{fp}, καθώς και του αριθμού των στοιχείων που πρόκειται να εισαχθούν στο φίλτρο, μπορούμε να επιλέξουμε το μέγεθος m του δυαδικού πίνακα να είναι τέτοιο που να εξασφαλίζει ότι τα λάθη δεν θα ξεπερνούν σε ποσοστό την πιθανότητα p_{fp} (βλέπε Παράρτημα). Το μέγεθος m δίνεται από τον παρακάτω τύπο

$$m = -\frac{n \cdot \ln(p)}{\ln(2)^2}$$ Εξ. 3.5

Μάλιστα, η πραγματική πιθανότητα false positive αναμένουμε να είναι πολύ μικρότερη από p_{fp}. Στο μοντέλο μας αρχικά, δημιουργούνται N^2 κλειδιά. Οι κόμβοι και οι επαφές τους αναπαρίστανται με φίλτρα Bloom στα οποία κατακερματίζονται τα παραπάνω κλειδιά. Συνεπώς, το μέγιστο πλήθος κλειδιών που μπορεί να περιέχει ένα φίλτρο είναι N^2. Η πιθανότητα p_{fp} αναφέρεται στο μέγιστο πλήθος στοιχείων του φίλτρου. Αυτή η περίπτωση, όμως, είναι ακραία και δεν αναμένεται να παρουσιάζεται στην πραγματικότητα. Σε αυτή την περίπτωση θα έπρεπε κάποιος κόμβος να έχει στη λίστα επαφών του όλους τους κόμβους του δικτύου. Κάτι τέτοιο, όμως, είναι πάρα πολύ δύσκολο να συμβεί στο είδος δικτύων που μελετάμε. Υποθέτοντας, λοιπόν, μια πιθανότητα p_{fp} και κατασκευάζοντας το φίλτρο με μέγιστο αριθμό στοιχείων N^2, αναμένουμε ότι η πιθανότητα λάθους θα είναι εν τέλει μικρότερη από p_{fp}, μιας και τα στοιχεία-κλειδιά δεν πρόκειται να φτάσουν στο μέγιστο δυνατό αριθμό που μπορεί να υποστηρίξει το φίλτρο. Το αντίκτυπο των false positives στην δρομολόγηση το μελετάμε πειραματικά στο επόμενο κεφάλαιο.

Μια άλλη παρατήρηση είναι ότι η απόφαση δρομολόγησης επηρεάζεται από τη σχετική τιμή των μετρικών. Ένας λάθος υπολογισμός του SimBetUtil δεν σημαίνει απαραίτητα λάθος απόφαση δρομολόγησης. Όπως περιγράψαμε στην ενότητα 3.3.4, ένας κόμβος αποφασίζει αν θα αναλάβει αυτός την προώθηση ενός πακέτου συγκρίνοντας τη δική του τιμή SimBetUtil με το SimBetUtil του άλλου κόμβου, με τον οποίο συνδιαλέγεται. Έστω π.χ. ένας κόμβος A ο οποίος έχει SimBetUtil (A, dst) = 0.9 και ένας κόμβος B έχει SimBetUtil (B, dst) = 0.5. Ο υπολογισμός των παραπάνω μετρικών έγινε χωρίς τη χρήση φίλτρων Bloom. Με βάση τον αλγόριθμο ο A, έχοντας μεγαλύτερη τιμή, θα αναλάβει την προώθηση του πακέτου. Έστω τώρα ότι οι παραπάνω μετρικές υπολογίζοντας με χρήση φίλτρων Bloom είναι οι παρακάτω: $SimBetUtil_{BF}$(A, dst) = 0.95 και $SimBetUtil_{BF}$(B, dst) = 0.8 δηλ. έχουμε λάθος λόγω της p_{fp} και στις δύο μετρικές. Σε αυτή την περίπτωση η απόφαση δρομολόγησης δεν θα αλλάξει, αφού οι νέες τιμές δεν διαφέρουν στη διάταξη τους από τις αρχικές. Δηλ. ο A έχει και σε αυτή τη περίπτωση μεγαλύτερο SimBetUtil από

τον B, οπότε αυτός θα προωθήσει το πακέτο. Επιπλέον, δεν πρέπει να ξεχνάμε ότι όσον αφορά την μετρική του betweenness, η πιθανότητα λάθους, όπως δείξαμε στην παράγραφο 3.3.2, είναι πολύ μικρή. Αυτό σημαίνει ότι στην πράξη, συνδυάζοντας τις μετρικές betweenness και similarity, η τιμή του SimBetUtil, που καθορίζει την απόφαση δρομολόγησης, δεν θα επηρεαστεί σε μεγάλο βαθμό.

Από τις παραπάνω παρατηρήσεις, βλέπουμε ότι τα λάθη στον υπολογισμό των μετρικών δεν συνεπάγονται πάντα μείωση του ποσοστού παράδοσης δεδομένων. Θέτοντας μικρό ποσοστό λάθους στην κατασκευή των φίλτρων σε συνδυασμό με τον τρόπο που χρησιμοποιούνται οι μετρικές στον SimBet και το πραγματικό ποσοστό λάθους στο δίκτυο, περιμένουμε πολύ λίγα λάθη στη δρομολόγηση. Άλλωστε μια λάθος απόφαση δρομολόγησης, δεν σημαίνει ότι το πακέτο έχει χαθεί, αλλά ότι μπορεί να φτάσει στον προορισμό από εναλλακτικό μονοπάτι, πιθανόν μέσω περισσότερων αλμάτων ή με μεγαλύτερη καθυστέρηση. Το αντίκτυπο της χρήσης φίλτρων Bloom στη γενικότερη επίδοση του δικτύου τη μελετάμε στο επόμενο κεφάλαιο.

3.6. Ζητήματα υλοποίησης

Εκτός από τα ζητήματα απόδοσης που αναφερθήκαν στην προηγούμενη παράγραφο, έχει νόημα να μελετήσουμε και κάποια θέματα υλοποίησης του αλγορίθμου που σχετίζονται με τη χρήση φίλτρων Bloom. Το βασικότερο ζήτημα υλοποίησης αφορά το μέγεθος του φίλτρου, δηλ. το μέγεθος του δυαδικού πίνακα που το υλοποιεί. Από την Εξ. 3.5 μπορούμε να υπολογίσουμε το μέγεθος του φίλτρου, δοθείσης της πιθανότητας p_{fp} και του πλήθους των στοιχείων που αναμένεται να εισαχθούν. Θεωρώντας ένα δίκτυο με 50 κόμβους και μια πιθανότητα $p_{fp} = 1\%$, δεδομένου ότι το πλήθος των κλειδιών είναι N^2, θα πρέπει να κατασκευάσουμε έναν πίνακα μεγέθους $m \approx 24000$ bits δηλ. 3000 bytes. Το μέγεθος των φίλτρων θα πρέπει να είναι ίδιο, ανεξάρτητα από το αν θα αναπαριστούν επαφές ή κόμβους πάνω στις κεφαλίδες των πακέτων. Αν συνυπολογίσουμε και τα επιπλέον κλειδιά που αναφέρονται σε εικονικούς κόμβους το παραπάνω μέγεθος μπορεί να είναι ακόμη μεγαλύτερο. Αυτό δημιουργεί ένα ζήτημα υλοποίησης, αλλά και επεκτασιμότητας. Για να αντιμετωπίσουμε το παραπάνω ζήτημα, υπάρχουν κάποιες προσεγγίσεις που

μπορούμε να εφαρμόσουμε και οι οποίες εκμεταλλεύονται τις ιδιότητες των δικτύων που μελετάμε.

Όπως έχουμε ήδη αναφέρει στην ενότητα 3.5, τα opportunistic δίκτυα είναι αραιά δηλ. κάθε κόμβος συναντά μικρό σχετικά αριθμό άλλων κόμβων, λόγω της ύπαρξης διαμερίσεων. Συνεπώς, δεν αναμένουμε από τους κόμβους να εξαντλήσουν όλο το χώρο των N^2 κλειδιών που μπορεί να αναπαραστήσει το φίλτρο. Αυτό μας επιτρέπει να κατασκευάζουμε το φίλτρο με βάση όχι το μέγιστο πλήθος κλειδιών, άλλα θεωρώντας ότι θα υπάρχουν κατά μέσο όρο πολύ λιγότερα από N^2 κλειδιά. Αυτό μας επιτρέπει να διατηρούμε την πιθανότητα p_{fp} ίδια, μειώνοντας το μέγεθος του φίλτρου. Σε αυτή την προσέγγιση τίθεται το ερώτημα ποιο είναι το ιδανικό μέγεθος του φίλτρου Bloom. Η απάντηση δεν είναι προφανής και είναι δύσκολο να καθοριστεί κάποιο σαφές όριο.

Μια άλλη λύση που μπορεί να εφαρμοστεί όταν το φίλτρο είναι αραιό, δηλ. περιέχει πολλά μηδενικά bit, είναι αντί να στέλνουμε όλο το φίλτρο, να αποστέλλονται μόνο οι θέσεις των άσσων. Όταν έχουμε ένα δίκτυο αραιό, όπου οι επαφές μεταξύ των κόμβων είναι λίγες, αυτό σημαίνει ότι τα φίλτρα θα έχουν πολύ λίγα bits ίσα με 1. Αντί να ακολουθήσουμε την προηγούμενη προσέγγιση, στην οποία θα πρέπει να προβλέψουμε τον αναμενόμενο αριθμό επαφών, μπορούμε να κατασκευάσουμε το φίλτρο με τον κλασικό τρόπο (υποστηρίζοντας N^2 κλειδιά). Κατά τη διαδικασία δρομολόγησης, όταν απαιτείται μεταφορά του φίλτρου, αντί να μεταφέρουμε όλη τη δομή, στέλνουμε μόνο τις θέσεις που έχουν τιμή 1. Έστω το παρακάτω παράδειγμα: θεωρούμε το φίλτρο του Πίνακα 3.2.

Πίνακας 3.2 Φίλτρο Bloom μεγέθους 32 bit

0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0

Το φίλτρο έχει μέγεθος 32 bit, άλλα μόνο τέσσερα από αυτά έχουν τεθεί μηδέν. Κάθε θέση του δυαδικού πίνακα μπορεί να αναπαρασταθεί με χρήση 5 bit (2^5=32). Άρα, αντί να στείλουμε όλο το φίλτρο (32 bit) μπορούμε να στείλουμε έναν πίνακα με τις

θέσεις που έχουν τιμή 1, όπως φαίνεται στον Πίνακα 3.3. Σε αυτή την περίπτωση απαιτούνται 4 θέσεις άσσων×$\log_2(32)$ bits = 20 bits.

Πίνακας 3.3 Εναλλακτική αναπαράσταση του φίλτρου του Πίνακα 3.1

4	11	23	24

Αυτή η προσέγγιση φαίνεται ότι είναι απαραίτητη για την αναπαράσταση του προορισμού πάνω στο πακέτο. Σε αυτή την περίπτωση το μέγεθος του φίλτρου, όπως είδαμε, είναι τέτοιο που δεν μπορεί να χωρέσει στην αντίστοιχη κεφαλίδα του πακέτου. Επίσης, το φίλτρο που αναπαριστά τον προορισμό θα έχει στις περισσότερες θέσεις του πίνακα τιμές μηδέν επειδή στην αναπαράσταση του προορισμού με ένα φίλτρο Bloom χρησιμοποιούμε μόνο ένα κλειδί. Η χρήση ενός κλειδιού σημαίνει ότι μόνο k θέσεις στον πίνακα θα έχουν τιμή 1, επειδή τόσο είναι το πλήθος των συναρτήσεων κατακερματισμού. Συνεπώς, πρέπει να εφαρμόσουμε αυτή την πρακτική για να είναι εφικτή η προσθήκη του προορισμού στο πακέτο.

Μια άλλη προσέγγιση είναι να επιτρέπουμε μεγαλύτερες τιμές για την πιθανότητα p_{fp} κατά την κατασκευή του φίλτρου. Από την Εξ. 3.5 βλέπουμε ότι κρατώντας σταθερό τον αριθμό των στοιχείων (N^2) και αυξάνοντας την p_{fp}, το μέγεθος m του φίλτρου θα μειωθεί. Βέβαια, μεγαλύτερες τιμές της p_{fp} σημαίνει ότι επιτρέπουμε περισσότερα λάθη κατά τον υπολογισμό των μετρικών. Όπως, όμως, έχουμε συζητήσει στην ενότητα 3.3, η p_{fp} αποτελεί ένα άνω φράγμα στην πιθανότητα λάθους. Ένα φίλτρο που προορίζεται για να χωρέσει N^2 κλειδιά είναι σχεδόν αδύνατο να γεμίσει κατά τη λειτουργία ενός opportunistic δικτύου. Επίσης, υπενθυμίζουμε ότι ένα λάθος στον υπολογισμό της μετρικής δεν συνεπάγεται διαφορετική απόφαση δρομολόγησης. Συνεπώς, η παραπάνω τακτική φαίνεται να αποτελεί μια καλή εναλλακτική για μείωση του μεγέθους του φίλτρου.

Τέλος, υπάρχει μια λύση η οποία είναι ουσιαστικά παραλλαγή του βασικού αλγορίθμου και η οποία φαίνεται να μας ωφελεί όσον αφορά στη μείωση του μεγέθους του φίλτρου. Η βασική ιδέα είναι να μην κρατάμε το σύνολο των επαφών που είχε ένας κόμβος v, όπως γίνεται στον παραδοσιακό αλγόριθμο, άλλα μόνο τις

πιο πρόσφατες. Ένας κόμβος v είναι πολύ πιθανόν να διαθέτει επαφές οι οποίες εμφανίστηκαν στο παρελθόν και οι οποίες δεν έχουν εμφανιστεί ξανά για μεγάλο χρονικό διάστημα. Τέτοιες επαφές δεν έχει νόημα να μένουν αποθηκευμένες, καθώς δεν αντικατοπτρίζουν την κοινωνική σχέση του v με αυτές. Αν εφαρμόσουμε την παραπάνω τακτική στην υλοποίηση του αλγορίθμου με φίλτρα Bloom, κάθε κόμβος θα κρατά λιγότερες επαφές στο φίλτρο του και θα το ανανεώνει ανά τακτά διαστήματα. Συνεπώς, δεν απαιτείται η κατασκευή μεγάλων φίλτρων που θα δυσκόλευαν την επικοινωνία.

Από τις παραπάνω προτάσεις γίνεται σαφές ότι το μέγεθος του φίλτρου, που φαίνεται να αποτελεί ανασταλτικό παράγοντα στην εφαρμογή της λύσης μας, μπορεί να περιοριστεί τόσο, ώστε τελικά να μπορεί ο αλγόριθμος να είναι λειτουργικός.

ΚΕΦΑΛΑΙΟ 4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Στο κεφάλαιο αυτό παρουσιάζουμε τα πειραματικά αποτελέσματα των προσομοιώσεων που πραγματοποιήθηκαν για την αξιολόγηση του αλγορίθμου. Αρχικά, περιγράφουμε το εργαλείο προσομοίωσης που χρησιμοποιήθηκε και τον τρόπο διεξαγωγής των πειραμάτων. Έπειτα, ακολουθούν οι μετρικές αξιολόγησης, τα πειράματα και ο σχολιασμός τους.

4.1. Περιβάλλον προσομοίωσης

Για την αξιολόγηση του αλγορίθμου που προτείνεται χρησιμοποιήσαμε τον προσομοιωτή *ONE* (Opportunistic Network Environment) [10]. Ο ONE είναι γραμμένος σε γλώσσα JAVA και σχεδιασμένος ειδικά για τον τύπο δικτύου που εξετάζουμε. Τα βασικότερο πλεονέκτημά του είναι ότι υπάρχει η δυνατότητα εισαγωγής επαφών από αρχείο. Αυτό σημαίνει ότι αντί να χρησιμοποιήσουμε κάποιο συνθετικό μοντέλο κίνησης των κόμβων (π.χ. Random Waypoint), μπορούμε να χρησιμοποιήσουμε καταγραφές επαφών μεταξύ κινούμενων χρηστών πραγματικών δικτύων. Τέτοιου είδους καταγραφές έχουν γίνει με χρήση φορητών συσκευών που μετέφεραν οι συμμετέχοντες στο πείραμα και τα δεδομένα των επαφών που έχουν εξαχθεί είναι διαθέσιμα στην επιστημονική κοινότητα, μέσω διαδικτύου [4]. Ο ONE μπορεί να διαβάσει τέτοιες επαφές καταγεγραμμένες σε απλό αρχείο κειμένου, με

κατάλληλη μορφοποίηση. Η χρήση πραγματικών επαφών, αντί κάποιου συνθετικού μοντέλου κατά τη διεξαγωγή των πειραμάτων, αποδίδει ρεαλιστικότερα τη λειτουργία ενός δικτύου και κατά συνέπεια οδηγεί σε ασφαλέστερα συμπεράσματα. Τα δεδομένα επαφών που χρησιμοποιήσαμε στα πειράματά μας προέρχονται από το [4]. Στον πίνακα 5.1 φαίνονται κάποια χαρακτηριστικά των δεδομένων που χρησιμοποιήθηκαν.

Πίνακας 4.1 Χαρακτηριστικά των συνόλων επαφών των πειραμάτων

Όνομα συνόλου επαφών	Infocom '05	Cambridge	Milano PMTR	Reality mining
Διάρκεια (μέρες)	3	11	19	246
Αριθμός συσκευών	41	54	44	97
Πλήθος επαφών	22.459	10.873	11.895	54.667

Τα σύνολα των δύο πρώτων στηλών είναι μέρος του προγράμματος *Haggle* (Haggle project) [6]. Όσον αφορά το σύνολο Reality mining [17], ο όγκος των επαφών που περιείχε ήταν πολύ μεγάλος (περιέχει δεδομένα που συλλέγονταν για περίπου εννέα μήνες), με αποτέλεσμα να είναι αδύνατη η χρήση του για προσομοίωση με τον ONE, λόγω περιορισμών μνήμης και ταχύτητας. Συνεπώς, επιλέξαμε να χρησιμοποιήσουμε ένα υποσύνολό του, που περιλαμβάνει τους μήνες με το μεγαλύτερο αριθμό επαφών (Οκτώβριο, Νοέμβριο και Δεκέμβριο).

Η παράμετρος που μεταβάλλεται στα πειράματα είναι η πιθανότητα false positive (p_{fp}). Όπως περιγράψαμε στο κεφάλαιο 3, η χρήση φίλτρων Bloom συνεπάγεται πιθανότητα λαθών κατά τον υπολογισμό των μετρικών. Τα λάθη αυτά οφείλονται στο γεγονός ότι η επερώτηση για την ύπαρξη ενός στοιχείου στο φίλτρο μπορεί να δώσει μια λανθασμένα θετική απάντηση. Αυτό έχει ως συνέπεια η πληροφορία που περιέχεται στα φίλτρα να μην παρέχει ακριβή υπολογισμό των μετρικών betweenness και similarity. Αυτό που θέλουμε να εξετάσουμε στα πειράματα είναι πόσο επηρεάζει η ύπαρξη λαθών τον υπολογισμό των μετρικών και κατ' επέκταση τις αποφάσεις δρομολόγησης. Με βάση τις παρατηρήσεις στην ενότητα 3.5 αναμένουμε ότι δεν θα έχουμε σοβαρό αντίκτυπο στην απόδοση. Οι τιμές της πιθανότητας p_{fp} που

χρησιμοποιήσαμε στα πειράματα είναι 1%, 5%, 10%, 20%. Καθεμιά από αυτές τις τιμές εκφράζει το μέγιστο ποσοστό false positive που μπορεί να υποστηρίξει το φίλτρο. Για παράδειγμα, όταν κατασκευάζουμε ένα φίλτρο με $p_{fp} = 10\%$ σημαίνει ότι επιλέγουμε το μέγεθος m του πίνακα που το αναπαριστά να είναι τέτοιο ώστε το ποσοστό των false positives να είναι το πολύ 10%, δηλαδή οι λανθασμένα θετικές απαντήσεις που θα λαμβάνουμε κατά τον έλεγχο ύπαρξης ενός στοιχείου (διεύθυνση ενός κόμβου) να μην ξεπερνούν το 10%. Η επιλογή του μεγέθους m καθορίζεται από την Εξ. 3.5. Οι παραπάνω τιμές αποτελούν τις ονομαστικές τιμές του ποσοστού false positives και εκφράζουν το άνω όριο που θέτουμε. Οι τιμές αυτές δεν ταυτίζονται με τις πραγματικές, αφού έχουν καθοριστεί με βάση το μέγιστο πλήθος στοιχείων που μπορούν να χωρέσουν σε ένα φίλτρο. Αυτό σημαίνει ότι για να παρουσιαστούν οι τιμές αυτές στην πραγματικότητα, θα πρέπει σε ένα φίλτρο επαφών ενός κόμβου να περιέχονται όλοι οι υπόλοιποι κόμβοι του δικτύου, κάτι που είναι σπάνιο να συμβεί σε ένα opportunistic δίκτυο.

Όσον αφορά την τηλεπικοινωνιακή κίνηση του δικτύου, ακολουθήσαμε το μοντέλο που χρησιμοποιείται συχνά στη βιβλιογραφία. Συγκεκριμένα, ο κάθε κόμβος του δικτύου παράγει ακριβώς ένα μήνυμα προς κάθε άλλο κόμβο. Τα μηνύματα αυτά παράγονται κατά την έναρξη της προσομοίωσης.

4.2. Μέθοδος Αξιολόγησης - Μετρικές

Για την αξιολόγηση της μεθόδου συγκρίναμε τις επιδόσεις του νέου αλγορίθμου σε σύγκριση με την παραδοσιακή έκδοση του αλγορίθμου SimBet. Η σύγκριση έγινε προσομοιώνοντας τους δύο αλγορίθμους σε καθένα από τα προαναφερθέντα σύνολα επαφών, για διάφορες τιμές της πιθανότητας p_{fp}. Οι μετρικές που λάβαμε υπ' όψιν ήταν οι παρακάτω:

Ποσοστό επιτυχούς παράδοσης πακέτων (Delivery Ratio): Η μετρική αυτή ορίζεται ως το πλήθος των πακέτων που παραδόθηκαν στον τελικό προορισμό τους προς το συνολικό πλήθος των πακέτων. Σε ένα τέτοιο δίκτυο, η αξιόπιστη μεταφορά δεδομένων δεν είναι εύκολο να επιτευχθεί. Οι απώλειες πακέτων είναι συνηθισμένες και οφείλονται σε φαινόμενα όπως οι διαμερίσεις ή οι περιορισμένοι πόροι (μνήμη) των κόμβων. Η μετρική αυτή, λοιπόν, αποτελεί μέτρο αξιοπιστίας του αλγόριθμου.

Όσο μεγαλύτερο είναι το ποσοστό αυτό, τόσο πιο ικανός είναι ο αλγόριθμος να παραδώσει ένα πακέτο στον προορισμό του. Στα πειράματα που πραγματοποιήσαμε καταγράψαμε τη μετρική τόσο συναρτήσει των false positives, όσο και σε συνάρτηση με το χρόνο.

Μέση καθυστέρηση (Average delay): Η μετρική αυτή είναι ο μέσος χρόνος που απαιτείται για να φτάσει ένα πακέτο στον προορισμό του. Από τη φύση τους τα opportunistic δίκτυα παρουσιάζουν αναπόφευκτα καθυστερήσεις στην παράδοση των πακέτων. Συνεπώς, οι αλγόριθμοι που επιτυγχάνουν μικρότερες καθυστερήσεις υπερέχουν.

Μέσος αριθμός αλμάτων (Average hops): Η μετρική αυτή ορίζεται ως το μέσο πλήθος αλμάτων που απαιτείται για να φτάσει ένα πακέτο στον προορισμό του, δηλ. από πόσους ενδιάμεσους κόμβους πέρασε μέχρι τον τελικό παραλήπτη. Η μετρική αυτή θέλουμε να είναι όσο γίνεται μικρότερη. Λιγότερα άλματα σημαίνει λιγότερες προωθήσεις πακέτων, άρα λιγότερη κατανάλωση ενέργειας. Η κατανάλωση ενέργειας είναι μια παράμετρος που δεν εξετάζουμε στην παρούσα διατριβή, αλλά αποτελεί σημαντικό ζήτημα για τις φορητές συσκευές.

Συνολικός αριθμός προωθήσεων (Total number of forwards): Η μετρική αυτή μας δείχνει πόσα πακέτα συνολικά προωθήθηκαν κατά τη διάρκεια της προσομοίωσης και σχετίζεται ως ένα βαθμό με την προηγούμενη μετρική (μέσος αριθμός αλμάτων). Περισσότερα άλματα συνεπάγονται περισσότερες προωθήσεις, αφού σε κάθε άλμα το πακέτο επαναπροωθείται. Ωστόσο, πρέπει να τονίζουμε ότι ο συνολικός αριθμός προωθήσεων αναφέρεται στο σύνολο των πακέτων που προωθούνται σε αντίθεση με το μέσο αριθμό αλμάτων που αναφέρεται μόνο σε πακέτα που έχουν παραδοθεί. Μικρότερες τιμές στον αριθμό των προωθήσεων σημαίνει λιγότερο τηλεπικοινωνιακό φορτίο για το δίκτυο και λιγότερες εκπομπές πακέτων. Συνεπώς, επιδιώκουμε όσο το δυνατόν χαμηλότερο αριθμό προωθήσεων.

Ποσοστό διαφορετικών αποφάσεων δρομολόγησης: Όπως έχουμε ήδη αναφέρει, η ύπαρξη false positives προκαλεί κάποια λάθη στον υπολογισμό των μετρικών. Οι λάθος υπολογισμοί, ωστόσο, δε συνεπάγονται απαραίτητα και διαφορετικές αποφάσεις δρομολόγησης. Η μετρική αυτή μετρά το πλήθος των διαφορετικών αποφάσεων δρομολόγησης σε σχέση με το συνολικό πλήθος αποφάσεων δρομολόγησης. Με τη χρήση αυτής της μετρικής δείχνουμε στη συνέχεια ότι η

πιθανότητα false positive δεν έχει ουσιαστική επίδραση στις αποφάσεις δρομολόγησης.

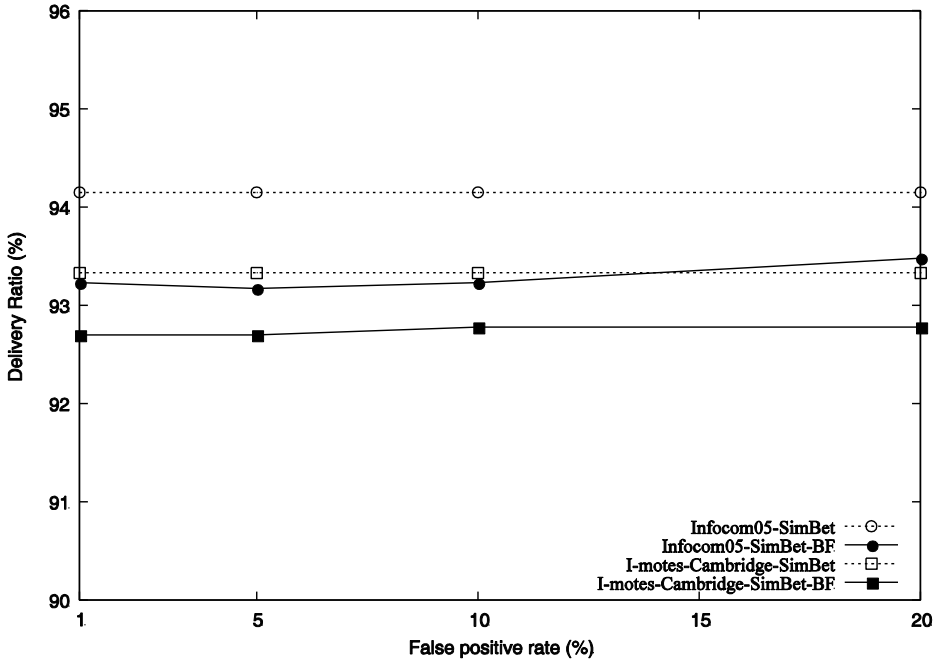
4.3. Πειράματα - Σχολιασμός Αποτελεσμάτων

Σε αυτή την ενότητα παραθέτουμε τα πειραματικά αποτελέσματα της προσομοίωσης με εκτενή σχολιασμό. Στα πειράματα συγκρίνουμε την επίδοση του SimBet-BF σε σχέση με τον SimBet γύρω από τρεις βασικούς άξονες: την ικανότητα του αλγορίθμου να παραδίδει τα πακέτα στους προορισμούς τους, την καθυστέρηση στην παράδοση των πακέτων και τον συνολικό αριθμό των εκπομπών. Τα αποτελέσματα που παρουσιάζονται στη συνέχεια ομαδοποιούνται με βάση τους παραπάνω άξονες.

4.3.1. Ικανότητα παράδοσης πακέτων

Στα σχήματα 4.1 και 4.2 βλέπουμε τα αποτελέσματα των πειραμάτων που αφορούν τη μετρική του ποσοστού επιτυχούς παράδοσης. Συγκεκριμένα, στο Σχήμα 4.1 φαίνονται τα αποτελέσματα για τα σύνολα επαφών Infocom05 και Cambridge, ενώ στο Σχήμα 4.2 για τα σύνολα επαφών Milano και Reality.

Παρατηρώντας τα αποτελέσματα, είναι εμφανές ότι το ποσοστό επιτυχούς παράδοσης πακέτων δεν έχει σημαντική διαφοροποίηση σε σχέση με τον αρχικό αλγόριθμο SimBet. Στο Infocom05 και στο Cambridge υπάρχει μια μικρή πτώση η οποία, όμως, δεν ξεπερνά το 1%. Η επίδοση του ανωνυμοποιημένου αλγορίθμου μάλιστα παραμένει σε αυτά τα επίπεδα ακόμα και όταν η p_{fp} τίθεται σε μεγάλες τιμές, φτάνοντας ακόμα και το 20%. Αυτή η σταθερότητα στην επίδοση οφείλεται κυρίως στο γεγονός ότι η p_{fp} είναι η ονομαστική τιμή του ποσοστού false positives και υπό την έννοια αυτή αποτελεί ένα άνω όριο για το ποσοστό των false positives. Η τιμή αυτή είναι δύσκολο να εμφανιστεί στην πραγματικότητα, επειδή, όπως έχουμε ήδη αναφέρει, αυτό θα σήμαινε ότι σε ένα φίλτρο που αναπαριστά τις επαφές ενός κόμβου περιέχονται όλα τα κλειδιά του δικτύου. Αυτό συνεπάγεται ότι κάποιος κόμβος θα είχε στη λίστα επαφών του όλους τους κόμβους του δικτύου, γεγονός που είναι σπάνιο.



Σχήμα 4.1 Ποσοστό επιτυχούς παράδοσης πακέτων συναρτήσει του false positive για τα σύνολα επαφών Infocom05 και Cambridge

Συνεπώς, τα λάθη στον υπολογισμό των μετρικών είναι πολύ μικρότερα από ότι η ονομαστική τιμή p_{fp}. Τα λιγότερα λάθη στους υπολογισμούς με τη σειρά τους συνεπάγονται λιγότερες διαφορές στις αποφάσεις δρομολόγησης, αφού οι δύο αλγόριθμοι θα υπολογίζουν ίδιες ή σχεδόν ίδιες τιμές των μετρικών betweenness και similarity. Το ποσοστό των διαφορετικών αποφάσεων δρομολόγησης για τα σύνολα Infocom05 και Cambridge φαίνεται στους πίνακες 4.2 και 4.3.

Πίνακας 4.2 Ποσοστό διαφορετικών αποφάσεων δρομολόγησης για το σύνολο επαφών Infocom05

Fp rate (%)	Different decisions (%)
1%	0.35%
5%	0.50%
10%	0.80%
20%	1.20%

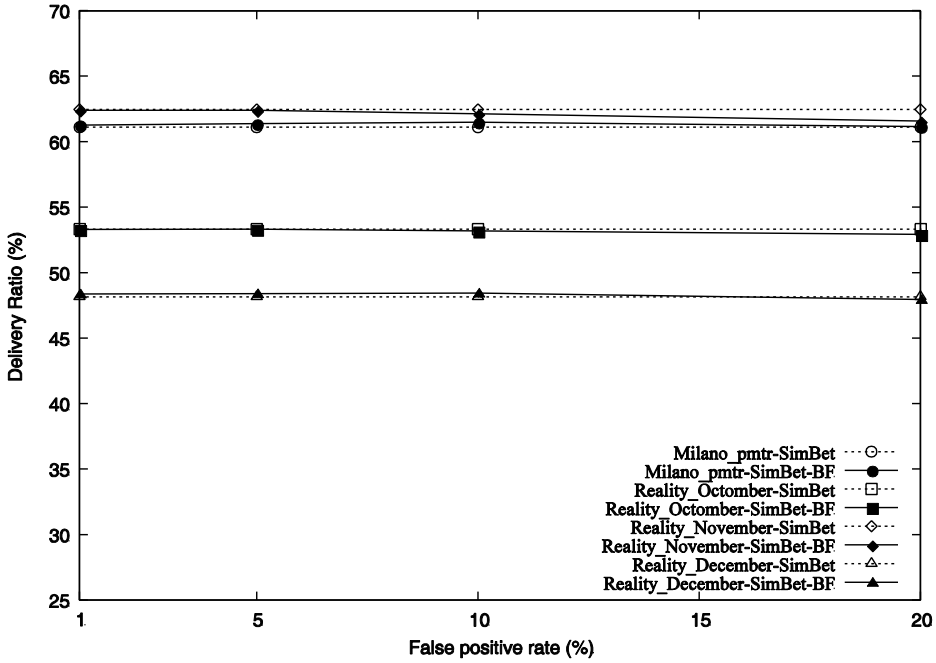
Παρατηρώντας τα αποτελέσματα στους πίνακες 4.2 και 4.3, βλέπουμε ότι το ποσοστό false positive έχει ελάχιστη επίδραση στις αποφάσεις δρομολόγησης. Το ποσοστό των

διαφορετικών αποφάσεων δρομολόγησης κυμαίνεται κάτω από 1% στις περισσότερες περιπτώσεις. Τα αποτελέσματα μπορούμε να πούμε ότι είναι αναμενόμενα και επιβεβαιώνουν τα όσα έχουν αναφερθεί στην ενότητα 3.5 σχετικά με την αποδοτικότητα του αλγορίθμου. Η ονομαστική τιμή p_{fp}, σε καμία περίπτωση δεν πλησιάζει το πραγματικό ποσοστό λαθών στη δρομολόγηση. Επίσης, είναι ο βασικός λόγος για τον οποίο δεν επηρεάζεται η αποδοτικότητα του SimBet-BF (Σχήμα 4.1).

Πίνακας 4.3 Ποσοστό διαφορετικών αποφάσεων δρομολόγησης για το σύνολο επαφών Cambridge

Fp rate (%)	Different decisions (%)
1%	0.32%
5%	0.32%
10%	0.45%
20%	0.90%

Ακόμη καλύτερη επίδοση παρατηρούμε στα σύνολα Milano και Reality (Σχήμα 4.2). Τα ποσοστό επιτυχούς παράδοσης είναι οριακά μικρότερο από τον αρχικό αλγόριθμο (της τάξης του 0,5%). Αυτή η πτώση, μάλιστα, εμφανίζεται μόνο όταν θέτουμε πολύ υψηλές τιμές p_{fp}. Σε μικρότερες τιμές της p_{fp} βλέπουμε ότι η μείωση είναι ακόμη μικρότερη. Το παραπάνω αποτέλεσμα είναι λογικό, αφού τα συγκεκριμένα σύνολα επαφών είναι πιο «αραιά» από τα προηγούμενα. Λέγοντας πιο «αραιά» εννοούμε ότι το πλήθος των επαφών ανά ημέρα είναι μικρότερο από τα προηγούμενα σύνολα επαφών. Αυτό φαίνεται ξεκάθαρα στον Πίνακα 4.1. Στο Infocom05 έχουμε, κατά μέσο όρο, περίπου 7.500 επαφές την ημέρα, ενώ στο Reality 222. Πιο αραιά σύνολα σημαίνει ότι λιγότερες επαφές εμπεριέχονται στο φίλτρο επαφών κάθε κόμβου. Επομένως, το πραγματικό ποσοστό false positives είναι σημαντικά μικρότερο από την ονομαστική τιμή p_{fp}. Επιπλέον, για μικρές τιμές p_{fp} (1% και 5%) παρατηρούμε ένα φαινόμενο το οποίο εκ πρώτης όψης μοιάζει παράδοξο. Σε κάποια από τα σύνολα π.χ. Milano ή Reality_October, η επίδοση του SimBet-BF ξεπερνά οριακά τον αρχικό αλγόριθμο. Λογικά, η επίδοση που θα περιμέναμε, για οποιοδήποτε σύνολο επαφών, θα έπρεπε να είναι πάντα μικρότερη ή το πολύ ίση με τον παραδοσιακό SimBet, λόγω της πιθανότητας p_{fp}.



Σχήμα 4.2 Ποσοστό επιτυχούς παράδοσης πακέτων συναρτήσει του false positive για τα σύνολα επαφών Milano και Reality

Η εξήγηση του παραπάνω αποτελέσματος βρίσκεται στον τρόπο κατασκευής των διευθύνσεων των κόμβων. Συγκεκριμένα, υπενθυμίζουμε ότι είναι δυνατή η υποεκτίμηση της τιμής similarity, κατά την αποστολή των επαφών ενός κόμβου Β σε έναν άλλον κόμβο Α (βλέπε υποενότητα 3.3.3). Το γεγονός αυτό φαίνεται ότι έχει θετική επίδραση στο ποσοστό επιτυχούς παράδοσης των πακέτων. Η διαφορετική απόφαση δρομολόγησης δεν σημαίνει απαραίτητα ότι το νέο μονοπάτι θα οδηγήσει σε απώλεια του πακέτου. Αντίθετα, αν το μονοπάτι που θα επέλεγε ο παραδοσιακός αλγόριθμος οδηγούσε σε αποτυχία, τότε ένα εναλλακτικό μονοπάτι (διαφορετική απόφαση δρομολόγησης) μπορεί να οδηγήσει σε επιτυχή παράδοση του πακέτου. Το παραπάνω σενάριο είναι πιο πιθανό να συμβεί σε σύνολα επαφών στα οποία η επίδοση του αλγορίθμου δεν είναι ιδιαίτερα υψηλή. Εκεί, ο βασικός αλγόριθμος κάνει αρκετές λανθασμένες επιλογές δρομολόγησης. Σε αυτή την περίπτωση, η διαφοροποίηση των αποφάσεων δρομολόγησης του SimBet-BF, μπορεί να αντισταθμίσουν τις λάθος επιλογές του αρχικού αλγορίθμου SimBet και το πακέτο να ακολουθήσει καλύτερη διαδρομή στο δίκτυο, η οποία θα οδηγήσει σε επιτυχή παράδοση.

Η διατήρηση του ποσοστού επιτυχούς παράδοσης σε υψηλά επίπεδα, όπως και στα προηγούμενα σύνολα επαφών, επιβεβαιώνεται και από το γεγονός ότι δεν έχουμε μεγάλη διαφορά στο ποσοστό διαφορετικών αποφάσεων δρομολόγησης. Στους πίνακες 4.5 - 4.8 φαίνονται τα ποσοστά αυτά για τα σύνολα Milano και Reality. Και πάλι είναι εμφανές ότι οι διαφορές κυμαίνονται κάτω από το 1% στις περισσότερες περιπτώσεις.

Πίνακας 4.4 Ποσοστό διαφορετικών αποφάσεων δρομολόγησης για το σύνολο επαφών Milano

Fp rate (%)	Different decisions (%)
1%	0.32%
5%	0.44%
10%	0.86%
20%	1.10%

Πίνακας 4.5 Ποσοστό διαφορετικών αποφάσεων δρομολόγησης για το σύνολο επαφών Reality_October

Fp rate (%)	Different decisions (%)
1%	0.12%
5%	0.18%
10%	0.38%
20%	0.99%

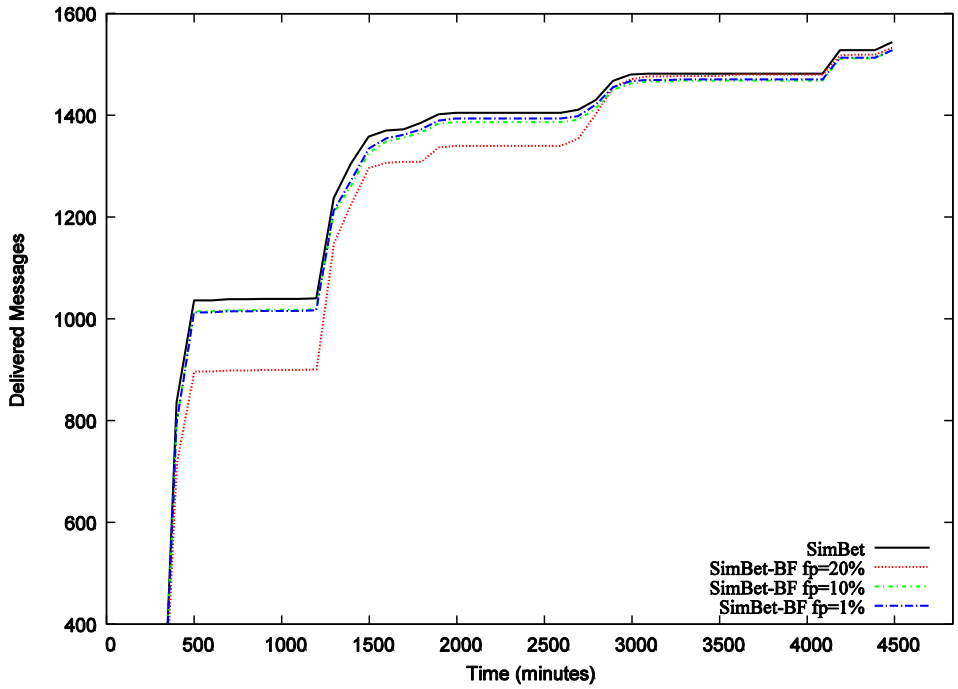
Πίνακας 4.6 Ποσοστό διαφορετικών αποφάσεων δρομολόγησης για το σύνολο επαφών Reality_November

Fp rate (%)	Different decisions (%)
1%	0.009%
5%	0.13%
10%	0.37%
20%	0.98%

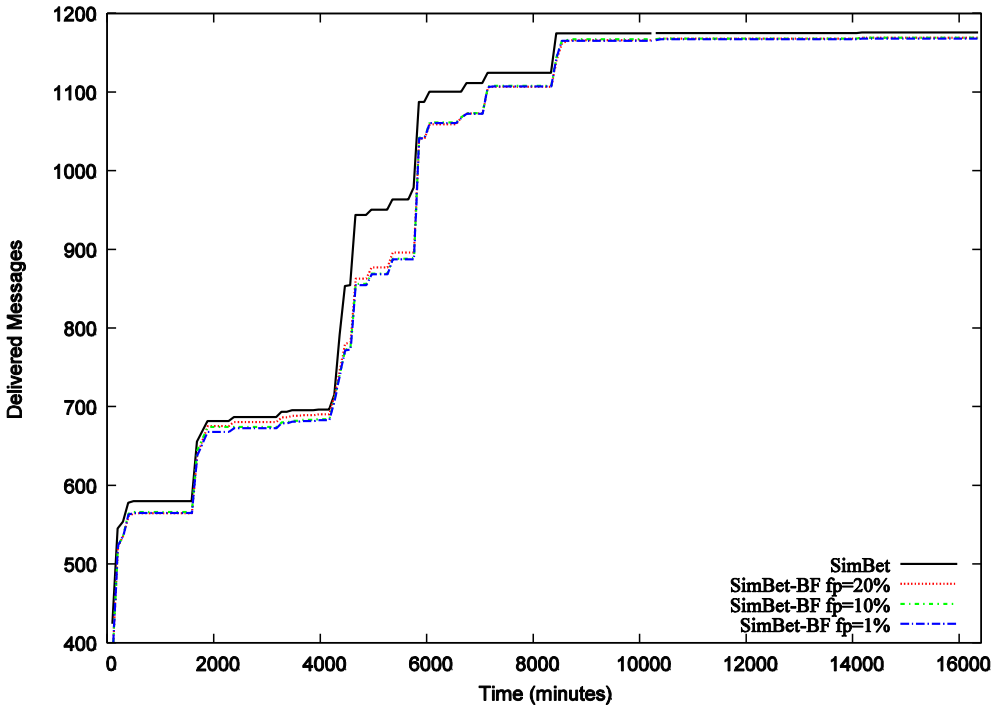
Πίνακας 4.7 Ποσοστό διαφορετικών αποφάσεων δρομολόγησης για το σύνολο επαφών Reality_December

Fp rate (%)	Different decisions (%)
1%	0.0089%
5%	0.11%
10%	0.22%
20%	0.75%

Στα Σχήματα 4.3 έως 4.8 βλέπουμε μια απεικόνιση των παραδιδόμενων πακέτων σε συνάρτηση με τον χρόνο. Η καταγραφή των πακέτων έγινε ανά 6000 δευτερόλεπτα (100 λεπτά). Είναι άξιο παρατήρησης το γεγονός ότι στα ενδιάμεσα σημεία ο αριθμός των πακέτων που έχουν παραδοθεί δεν είναι πάντα ίδιος με τον αρχικό αλγόριθμο. Αυτό οφείλεται στις διαφορετικές αποφάσεις δρομολόγησης που ενδέχεται να οδηγήσουν σε εναλλακτικές διαδρομές. Οι διαφορές αυτές είναι μεγαλύτερες όταν η ονομαστική τιμή p_{fp} λαμβάνει μεγάλες τιμές (π.χ. 20%).



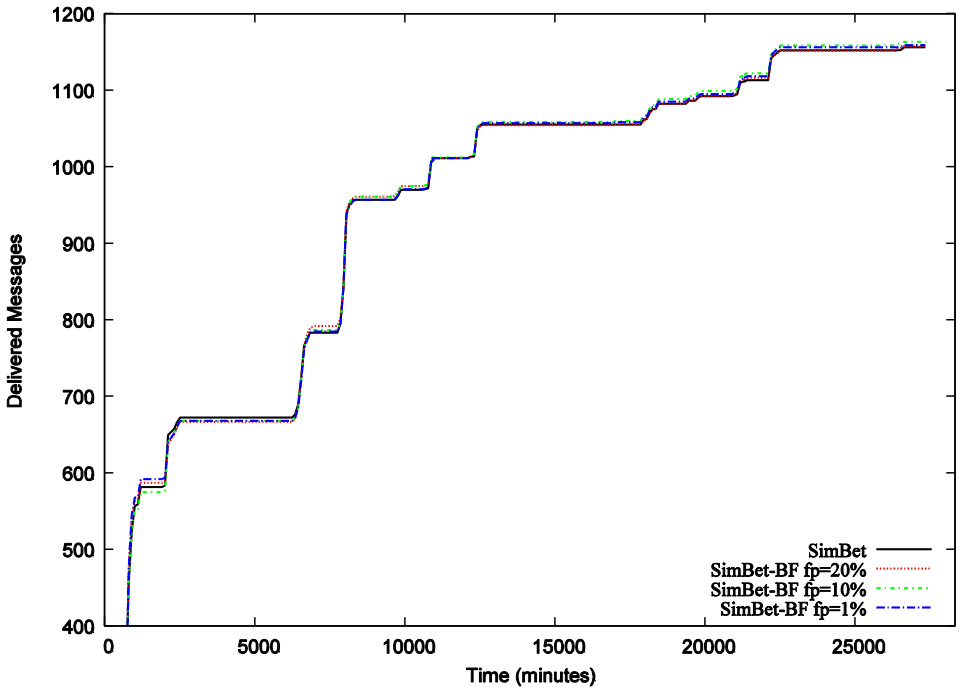
Σχήμα 4.3 Αριθμός παραδιδόμενων πακέτων συναρτήσει του χρόνου στο σύνολο επαφών Infocom05 για διάφορες τιμές p_{fp}



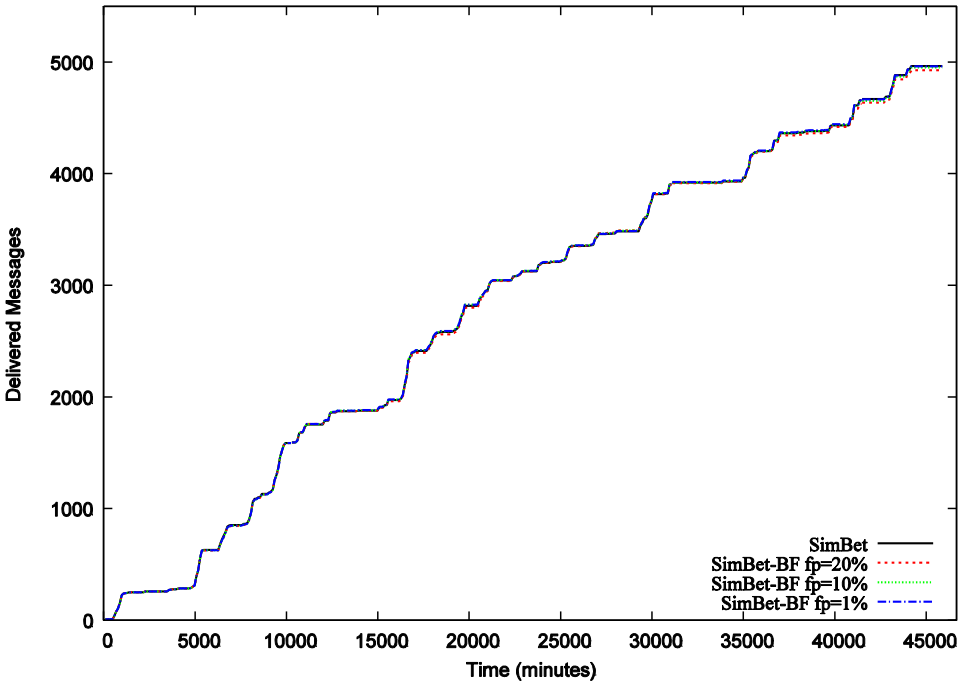
Σχήμα 4.4 Αριθμός παραδιδόμενων πακέτων συναρτήσει του χρόνου στο σύνολο επαφών Cambridge για διάφορες τιμές p_{fp}

Μεγάλες τιμές της πιθανότητας p_{fp} σημαίνουν περισσότερα λάθη στους υπολογισμούς των μετρικών, άρα μεγαλύτερη πιθανότητα διαφορετικών αποφάσεων δρομολόγησης, όπως είδαμε στους πίνακες 4.2 - 4.8. Οι διαδρομές αυτές περιλαμβάνουν περισσότερα άλματα ή εισάγουν μεγαλύτερη καθυστέρηση, χωρίς αυτό να ισχύει πάντα, όπως θα δούμε στην επόμενη ενότητα. Αυτή η διαφοροποίηση δεν σχετίζεται με το συνολικό αριθμό παραδιδόμενων πακέτων που είδαμε πριν, αλλά δείχνει πως οι διαφορετικές αποφάσεις δρομολόγησης στα ενδιάμεσα βήματα επηρεάζουν τη διαδρομή που ακολουθούν τα πακέτα στο δίκτυο.

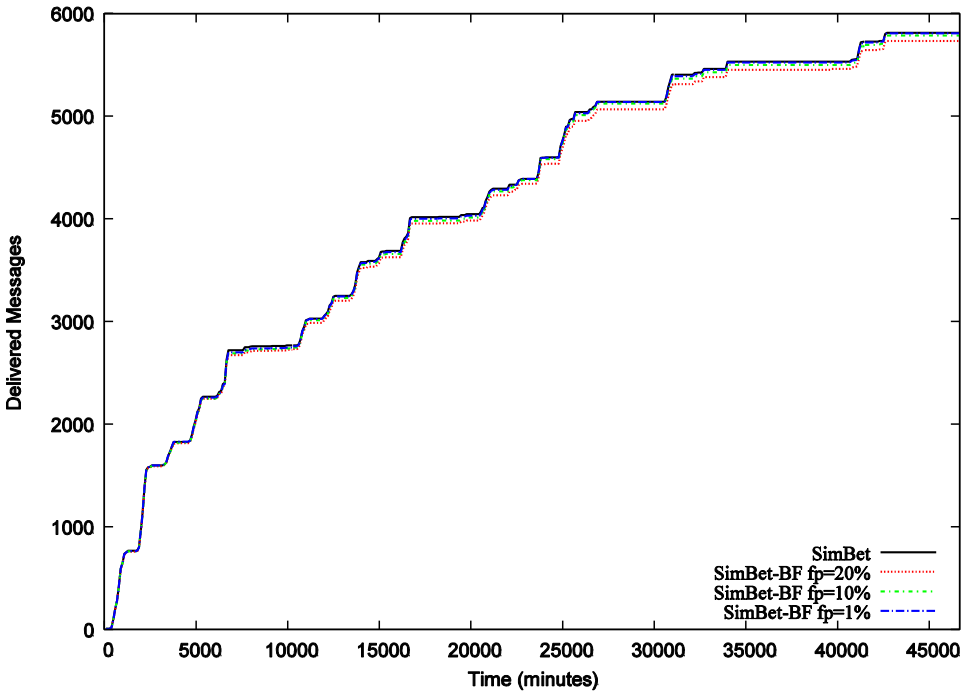
Οι διαφορές αυτές είναι ελάχιστες στα πιο αραιά σύνολα, όπως το Reality (Σχήματα 4.6 - 4.8). Εκεί, οι δύο αλγόριθμοι σχεδόν ταυτίζονται, αφού οι διαφορές στις αποφάσεις δρομολόγησης είναι λιγότερες (Πίνακες 4.6 - 4.8). Αυτό σημαίνει ότι οι διαδρομές που ακολουθούν τα πακέτα στο δίκτυο είναι σχεδόν ταυτόσημες με τον αρχικό αλγόριθμο.



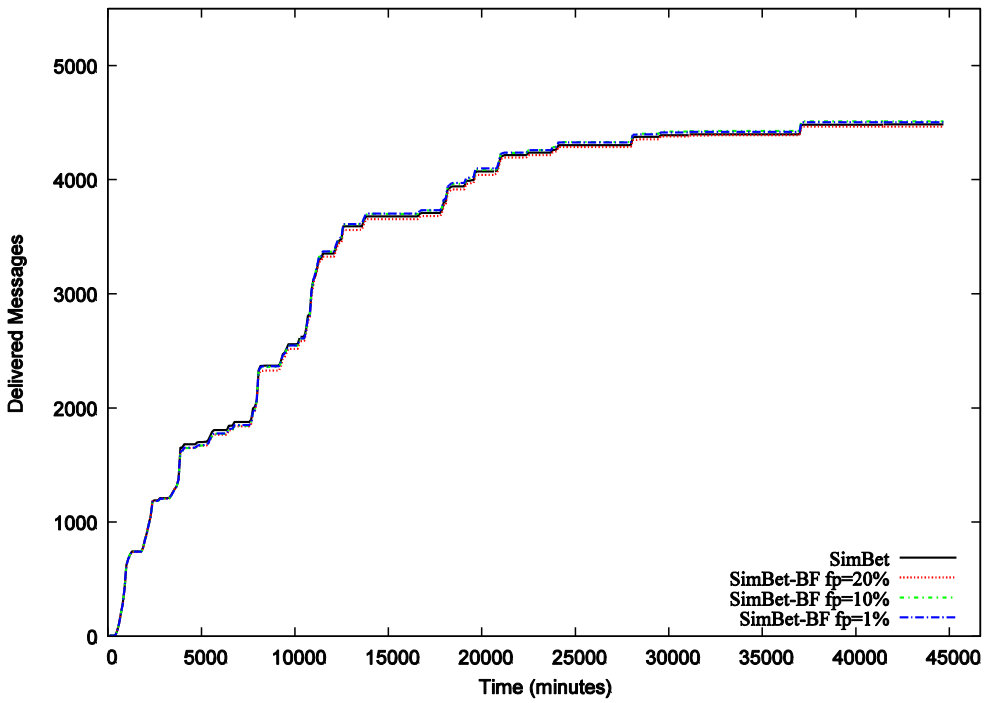
Σχήμα 4.5 Αριθμός παραδιδόμενων πακέτων συναρτήσει του χρόνου στο σύνολο επαφών Milano για διάφορες τιμές p_{fp}



Σχήμα 4.6 Αριθμός παραδιδόμενων πακέτων συναρτήσει του χρόνου στο σύνολο επαφών Reality_October για διάφορες τιμές p_{fp}



Σχήμα 4.7 Αριθμός παραδιδόμενων πακέτων συναρτήσει του χρόνου στο σύνολο επαφών Reality_November για διάφορες τιμές p_{fp}



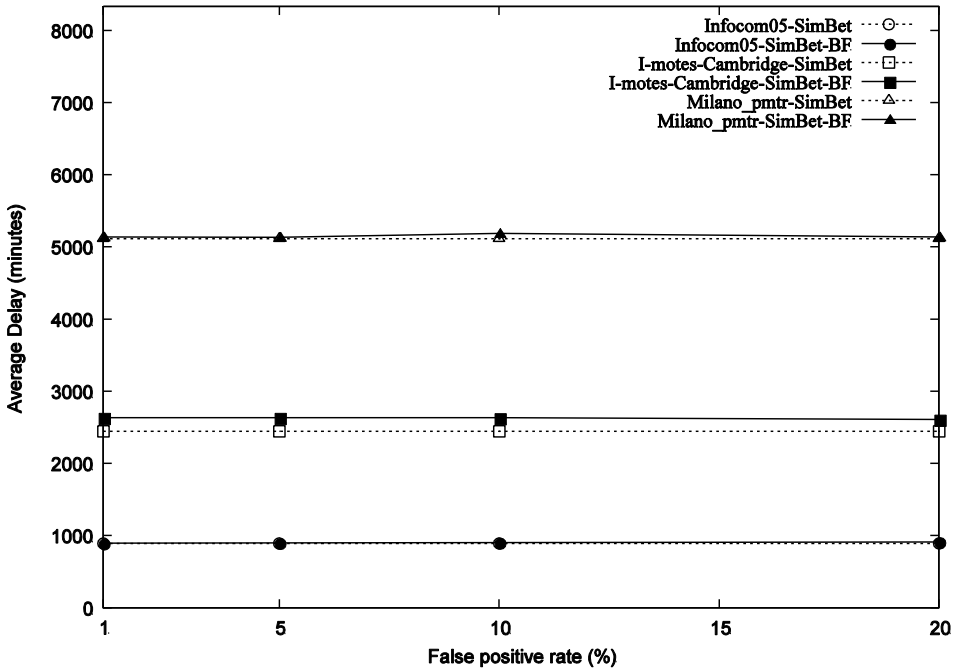
Σχήμα 4.8 Αριθμός παραδιδόμενων πακέτων συναρτήσει του χρόνου στο σύνολο επαφών Reality_December για διάφορες τιμές p_{fp}

Με βάση τα παραπάνω αποτελέσματα μπορούμε να πούμε ότι η ικανότητα του αλγόριθμου να παραδίδει επιτυχώς τα πακέτα στους προορισμούς δεν επηρεάζεται ουσιαστικά από τα λάθη στους υπολογισμούς των μετρικών που οφείλονται στη χρήση φίλτρων Bloom. Αυτό είναι αναμενόμενο αφού τα φίλτρα που αναπαριστούν τις επαφές δεν πρόκειται ποτέ να περιέχουν το μέγιστο πλήθος στοιχείων που μπορούν να υποστηρίξουν. Επομένως, τα λάθη κατά τη χρήση του φίλτρου των επαφών είναι πολύ λιγότερα απ' ότι έχει καθοριστεί κατά την κατασκευή του. Επίσης, ακόμη και όταν ένα φίλτρο δώσει λάθος απάντηση σχετικά με την ύπαρξη ενός στοιχείου, αυτό δεν συνεπάγεται λάθος στην απόφαση δρομολόγησης. Προφανώς, ο υπολογισμός μιας μετρικής δεν θα είναι πάντα ακριβής αλλά αυτό που καθορίζει την απόφαση δρομολόγησης είναι το αποτέλεσμα της σύγκρισης των μετρικών δύο κόμβων κατά την εκτέλεση του πρωτοκόλλου. Αν η σύγκριση δίνει ίδιο αποτέλεσμα με τον παραδοσιακό αλγόριθμο τότε η απόφαση δρομολόγησης θα είναι η ίδια. Δηλαδή αυτό που εν τέλει καθορίζει την απόφαση δρομολόγησης είναι η σχετική τιμή των μετρικών και όχι η ακριβής.

4.3.2. Καθυστέρηση παράδοσης πακέτων

Σε αυτή την υποενότητα εξετάζουμε την καθυστέρηση που εισάγει ο SimBet-BF στην παράδοση των πακέτων σε σχέση με τον SimBet. Στο σχήμα 4.9 εμφανίζεται η μέση καθυστέρηση για τα σύνολα επαφών Infocom05, Cambridge και Milano, ενώ στο Σχήμα 4.11 παρουσιάζονται τα αποτελέσματα για τους τρεις μήνες του Reality.

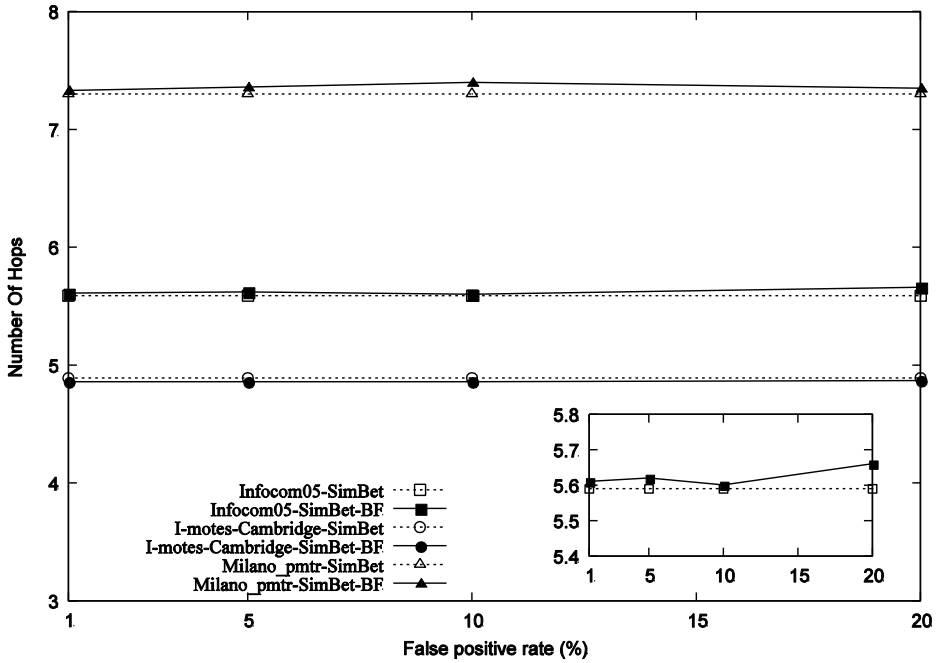
Εδώ παρατηρούμε ότι η καθυστέρηση ακολουθεί μια γενικότερη αυξητική τάση. Δεδομένων των λαθών στη δρομολόγηση, λόγω της λειτουργίας των φίλτρων Bloom, είναι αναμενόμενο τα πακέτα που δρομολογούνται λάθος να αργούν περισσότερο να φτάσουν στον προορισμό τους. Άλλωστε, όπως είδαμε στα διαγράμματα της προηγούμενης ενότητας (Σχήματα 4.3 - 4.8), ο αριθμός των παραδιδόμενων πακέτων στα ενδιάμεσα χρονικά διαστήματα δεν είναι ίδιος για τους δύο αλγορίθμους. Αυτό σημαίνει ότι πακέτα που δεν έχουν παραδοθεί σε κάποια χρονική στιγμή, θα φτάσουν στον προορισμό τους με κάποια καθυστέρηση σε σχέση με τον αρχικό αλγόριθμο. Η αύξηση της καθυστέρησης, στα περισσότερα σύνολα επαφών, δεν είναι μεγάλη και μπορούμε να πούμε με ασφάλεια ότι είναι ανεκτή για το δίκτυο.



Σχήμα 4.9 Μέση καθυστέρηση συναρτήσει του false positive για τα σύνολα επαφών Infocom05, Cambridge και Milano

Εκτός από το σύνολο επαφών Cambridge, όπου η διαφορά κυμαίνεται μεταξύ 6,5% και 8,5%, η αύξηση για τα υπόλοιπα σύνολα επαφών δεν ξεπερνά το 2-2,5%. Η παραπάνω αύξηση της καθυστέρησης στο συγκεκριμένο σύνολο επαφών είναι αναμενόμενη αν παρατηρήσουμε τα αντίστοιχα αποτελέσματα για τη μετρική του μέσου αριθμού αλμάτων (Σχήμα 4.10). Στο Σχήμα 4.10, για το σύνολο Cambridge, ο μέσος αριθμός αλμάτων είναι ελαφρώς μικρότερος από τον SimBet. Το παραπάνω σημαίνει ότι τα πακέτα, αφού δεν προωθούνται, μένουν περισσότερο χρόνο αποθηκευμένα στους κόμβους, αυξάνοντας την καθυστέρηση παράδοσης. Αυτό πιστεύουμε ότι οφείλεται στην ειδική περίπτωση υποεκτίμησης της μετρικής similarity (βλέπε ενότητα 3.3.3), η οποία μειώνει τις πιθανότητες ένα πακέτο να προωθηθεί σε επόμενο κόμβο. Υπενθυμίζουμε ότι η περίπτωση που ένας κόμβος (έστω A) υποεκτιμά τη μετρική similarity εμφανίζεται όταν έχει λάβει το summary vector από έναν γείτονά του (έστω B) και το επόμενο βήμα είναι να του ζητήσει μηνύματα για τους προορισμούς για τους οποίους διαθέτει μεγαλύτερη τιμή SimBetUtil. Η υποεκτίμηση της μετρικής similarity στον A μειώνει την πιθανότητα να έχει μεγαλύτερη τιμή SimBetUtil από τον B, άρα και την πιθανότητα να ζητήσει

από το γείτονά του να του προωθήσει πακέτα. Αυτό σημαίνει ότι, λόγω της υποεκτίμησης της μετρικής similarity, είναι πιθανόν οι κόμβοι, συχνά, να μην προωθούν τα πακέτα σε άλλους κόμβους.

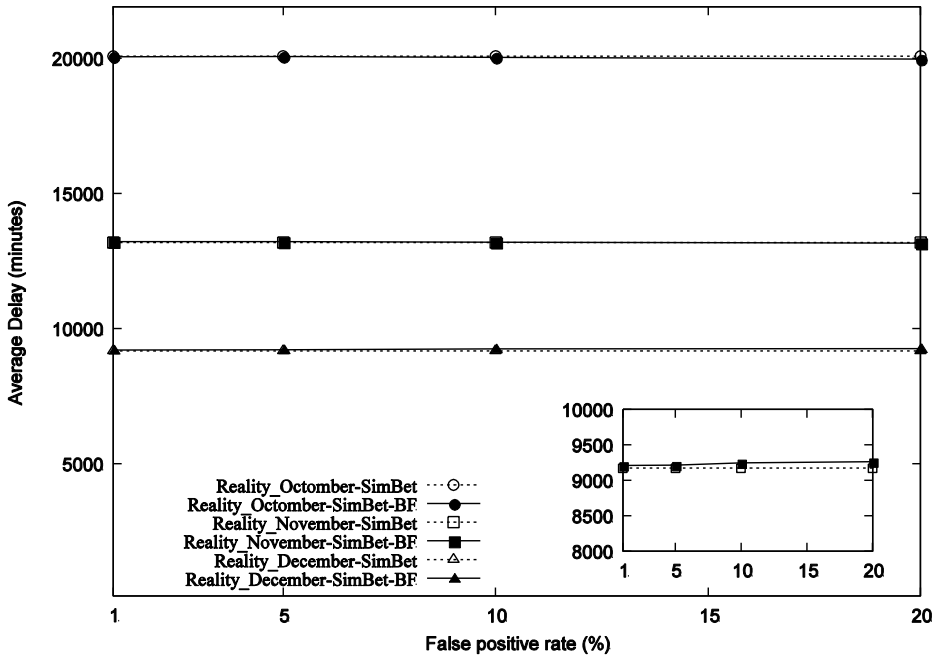


Σχήμα 4.10 Μέσος αριθμός αλμάτων συναρτήσει του false positive για τα σύνολα επαφών Infocom05, Cambridge και Milano

Πάντως, μπορούμε να πούμε ότι η παραπάνω περίπτωση είναι η εξαίρεση καθώς τα πακέτα, λόγω του διαφορετικού μηχανισμού δρομολόγησης, τείνουν να ακολουθούν δρομολόγια με μεγαλύτερο αριθμό αλμάτων. Η διαφορά αυτή, παρόλα αυτά, είναι πάρα πολύ μικρή σε σχέση με τον SimBet. Παρατηρώντας τα υπόλοιπα σύνολα επαφών, βλέπουμε ότι η αύξηση της μέσης καθυστέρησης ποτέ σχεδόν δεν ξεπερνά το 1%. Αυτό δικαιολογεί και τις μικρές διαφορές στην καθυστέρηση που είδαμε πριν, στο Σχήμα 4.9.

Είναι, επίσης, αξιοπρόσεκτο το γεγονός (όπως και στις προηγούμενες μετρικές) ότι η μέση καθυστέρηση δεν αυξάνεται αναλογικά με το ποσοστό των false positives, άλλα παρουσιάζει μια αξιοσημείωτη σταθερότητα. Αυτό αποτελεί άλλη μια πειραματική εξακρίβωση του γεγονότος ότι το πραγματικό ποσοστό των λαθών δρομολόγησης, απέχει πολύ από την ονομαστική τιμή p_{fp}.

Η παραπάνω διαφορά των SimBet και SimBet-BF ως προς την καθυστέρηση σχεδόν εκμηδενίζεται στο σύνολο επαφών Reality. Όπως έχουμε ήδη αναφέρει, σε αυτά τα σύνολα οι επαφές μεταξύ των κόμβων είναι πιο αραιές και τα φίλτρα ποτέ δεν φτάνουν το μέγιστο αριθμό επαφών που μπορούν να χωρέσουν.

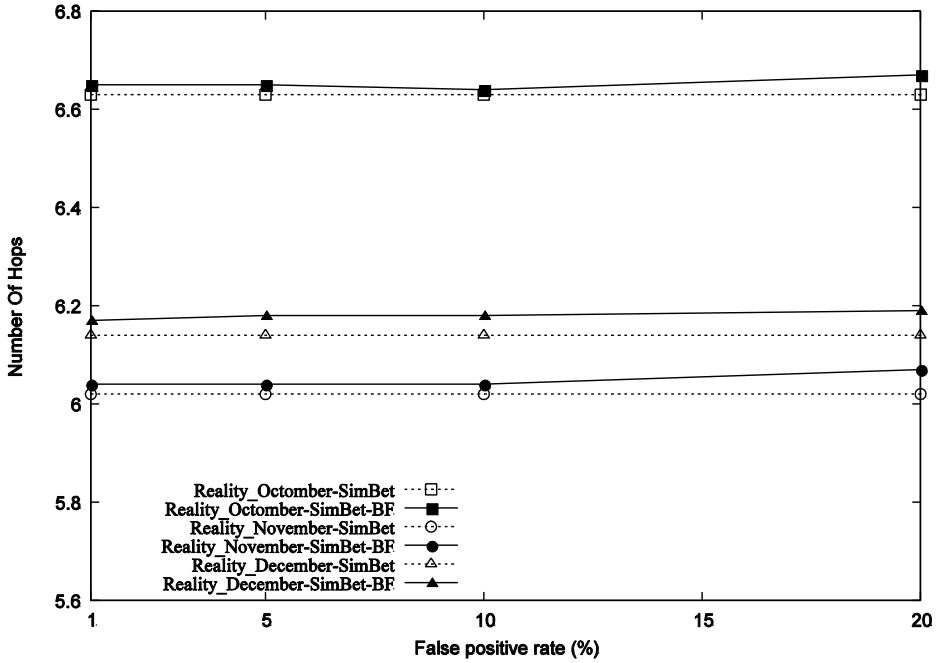


Σχήμα 4.11 Μέση καθυστέρηση συναρτήσει του false positive για το σύνολο επαφών Reality

Έτσι, τα λάθη στον υπολογισμό των μετρικών betweenness και similarity και η επιλογή διαφορετικών μονοπατιών συμβαίνει σε λιγότερες περιπτώσεις. Μάλιστα, παρατηρώντας το υποσύνολο των επαφών του μήνα Νοεμβρίου, βλέπουμε ότι η καθυστέρηση παρουσιάζει μια πολύ ελαφριά μείωση αντί για αύξηση. Αυτό έρχεται σε αντίθεση με τα μέχρι τώρα παρατηρούμενα αποτελέσματα και ουσιαστικά υποδηλώνει ότι το διαφορετικό μονοπάτι που επιλέγεται, ανεξαρτήτως αν περιλαμβάνει περισσότερους ενδιάμεσους κόμβους, οδηγεί εν τέλει σε γρηγορότερη παράδοση του πακέτου.

Στο Σχήμα 4.12 παρουσιάζεται ο μέσος αριθμός αλμάτων για τους τρεις μήνες του Reality. Τα αποτελέσματα δεν έχουν κάποια ουσιαστική διαφοροποίηση σε σχέση με τα προηγούμενα σύνολα επαφών. Η διαφοροποίηση του SimBet-BF σε σχέση με τον

βασικό αλγόριθμο είναι και πάλι ελάχιστη. Ο αλγόριθμος SimBet-BF κινείται σε ανάλογα επίπεδα με τον SimBet όσον αφορά το μέσο αριθμό αλμάτων (διαφορές της τάξης 0,6% - 0,8%). Αυτό επιβεβαιώνει τις μηδαμινές διαφορές των δύο αλγορίθμων ως προς την καθυστέρηση παράδοσης που φαίνονται στο Σχήμα 4.11. Όπως και στις προηγούμενες μετρικές, βλέπουμε ότι το ποσοστό των false positives δεν επηρεάζει ουσιαστικά τη σωστή λειτουργία του αλγορίθμου.

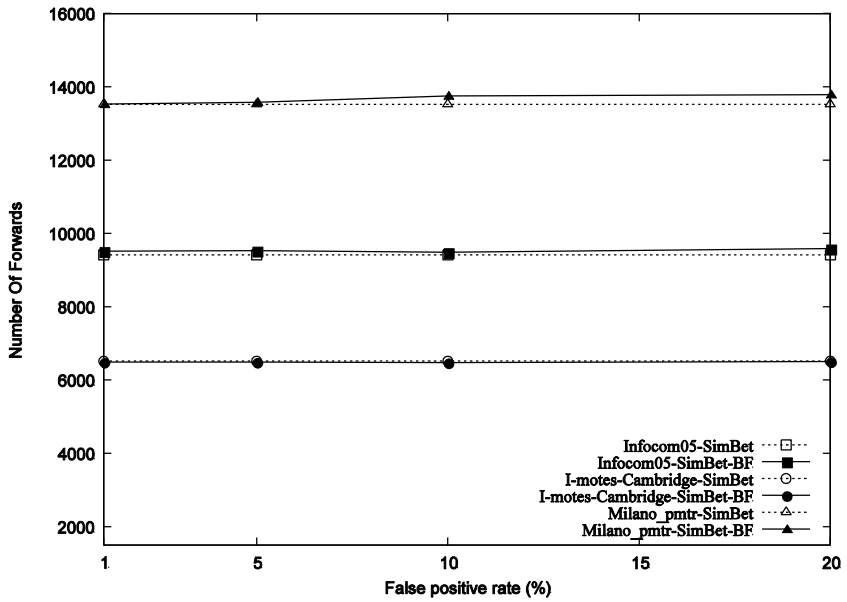


Σχήμα 4.12 Μέσος αριθμός αλμάτων συναρτήσει του false positive για το σύνολο επαφών Reality

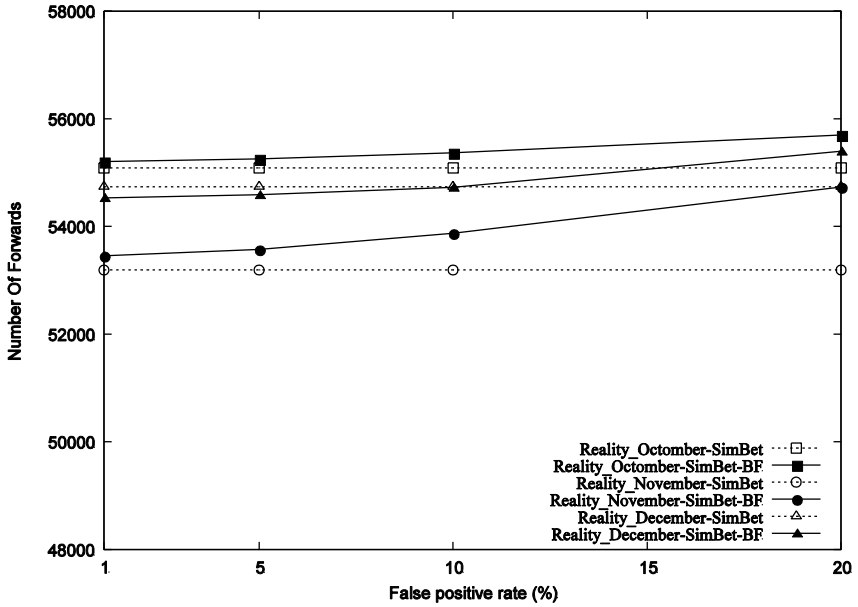
4.3.3. Συνολικός αριθμός εκπομπών

Στα σχήματα 4.13 και 4.14 παρουσιάζεται ο συνολικός αριθμός προωθήσεων των πακέτων συναρτήσει του ποσοστού p_{fp}. Επισημαίνουμε ότι προώθηση ενός πακέτου σημαίνει ότι το πακέτο επανεκπέμπεται.

Ο συνολικός αριθμός προωθήσεων στον SimBet-BF είναι αυξημένος σε σχέση με τον SimBet. Αυτό είναι λογικό ως ένα βαθμό δεδομένου ότι και ο αριθμός των αλμάτων, όπως είδαμε στην προηγούμενη ενότητα, ήταν ελαφρώς αυξημένος σχεδόν σε όλα τα σύνολα που εξετάζουμε (Σχήματα 4.10 και 4.12).



Σχήμα 4.13 Συνολικός αριθμός προωθήσεων συναρτήσει του false positive για τα σύνολα επαφών Infocom05, Cambridge και Milano



Σχήμα 4.14 Συνολικός αριθμός προωθήσεων συναρτήσει του false positive για το σύνολο επαφών Reality

Περισσότερα άλματα σημαίνει και περισσότερες προωθήσεις. Ωστόσο, ο αριθμός των προωθήσεων αναφέρεται στο σύνολο των πακέτων που διακινούνται στο δίκτυο, ενώ το πλήθος των αλμάτων μόνο στα πακέτα που παραδόθηκαν επιτυχώς. Συνεπώς, η αυξητική τάση δεν είναι ανάλογη με την αύξηση στα άλματα. Ο αριθμός των

προωθήσεων παρουσιάζει μια ελαφρώς μεγαλύτερη τάση αύξησης, καθώς μετράμε και τις προωθήσεις πακέτων τα οποία ποτέ δεν έφτασαν στον προορισμό τους. Ωστόσο, και πάλι η αύξηση αυτή δεν ξεπερνά το 2,8% (σύνολο επαφών Reality_November) σε σχέση με την παραδοσιακή έκδοση του SimBet. Ο λόγος που συμβαίνει αυτό είναι επειδή ο SimBet-BF «επιλέγει» να προωθεί περισσότερες φορές τα πακέτα προκαλώντας έτσι περισσότερες επανεκπομπές.

Επιπλέον, σε αυτές τις μετρήσεις παρατηρούμε ότι και πάλι υπάρχουν κάποιες περιπτώσεις όπου ο SimBet-BF παρουσιάζει καλύτερη επίδοση από τον SimBet. Για παράδειγμα, στο σύνολο Cambridge παρατηρούμε ότι ο αριθμός των προωθήσεων είναι ελαφρώς μειωμένος σε σχέση με τον αρχικό αλγόριθμο. Αυτό σχετίζεται με το μέσο αριθμό αλμάτων που είδαμε στην προηγούμενη ενότητα και οφείλεται στο γεγονός ότι μπορεί να έχουμε περιπτώσεις υποεκτίμησης της μετρικής similarity. Αν προσέξουμε στο διάγραμμα του Σχήματος 4.10, το ίδιο ισχύει και για τον αριθμό των αλμάτων για το συγκεκριμένο σύνολο. Αυτό σημαίνει ότι τα πακέτα διακινούνταν μέσω μικρότερου αριθμού ενδιάμεσων κόμβων, άρα ο αριθμός των επανεκπομπών είναι μικρότερος.

4.4. Λειτουργία και αξιολόγηση χωρίς τη χρήση εικονικών κόμβων

Στην ενότητα αυτή εξετάζουμε την απόδοση μιας εναλλακτικής προσέγγισης στο πρόβλημα προστασίας της ανωνυμίας που παρουσιάσαμε στις προηγούμενες ενότητες. Υπενθυμίζουμε ότι η ιδέα που παρουσιάσαμε στηρίζεται στην αντικατάσταση των διευθύνσεων των κόμβων με φίλτρα Bloom τα οποία περιέχουν ως στοιχεία τα κλειδιά που αντιστοιχίζονται στα ζεύγη των κόμβων από μια αξιόπιστη αρχή (TA). Εκτός από τις έγκυρες διευθύνσεις, για να εξασφαλιστεί η ανωνυμία κατά την επικοινωνία μεταξύ των κόμβων, παρέχονται επιπλέον διευθύνσεις στους κόμβους που αντιστοιχούν σε εικονικούς κόμβους, οι οποίοι δεν υπάρχουν στο δίκτυο (βλέπε ενότητα 3.3). Αυτό είναι απαραίτητο για τη διατήρηση της ανωνυμίας τόσο κατά την άμεση επικοινωνία όσο και κατά τη χρήση των επαφών ενός γείτονα για εξαγωγή των μετρικών betweenness και similarity (βλέπε ενότητα 3.4). Όπως είδαμε στην ενότητα 3.6 αυτή η προσέγγιση δημιουργεί ζητήματα υλοποίησης και επεκτασιμότητας. Ο λόγος που συμβαίνει αυτό είναι ότι το μέγεθος

του φίλτρου επαφών γίνεται μεγάλο. Συνεπώς, δημιουργείται ζήτημα κατά την αποστολή των επαφών μεταξύ των κόμβων.

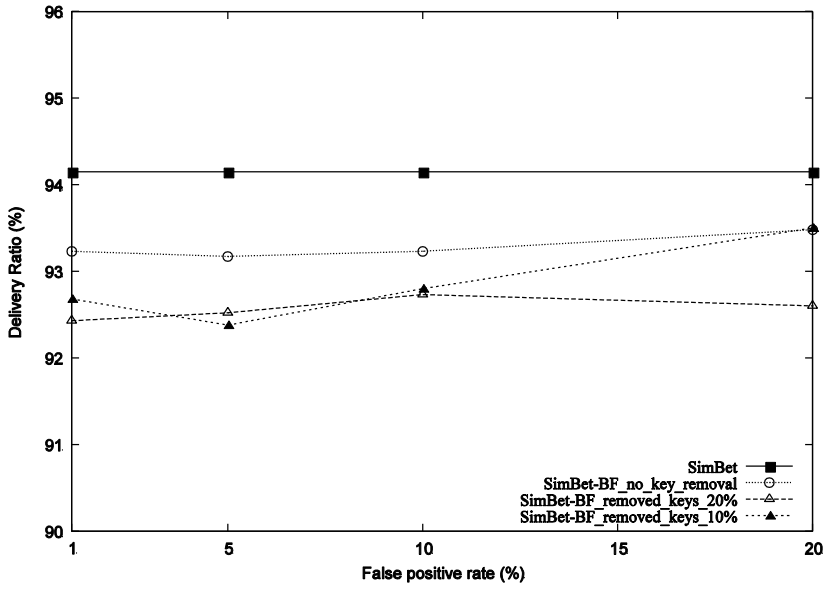
Για το λόγο αυτό παρέχουμε μια εναλλακτική προσέγγιση με στόχο να μειώσουμε το μέγεθος του φίλτρου. Η ιδέα είναι να μην χρησιμοποιήσουμε επιπλέον εικονικούς κόμβους. Επομένως, η ΤΑ δεν χρειάζεται να παράγει και να ενσωματώσει στις διευθύνσεις των κόμβων επιπλέον κλειδιά. Έτσι, το πλήθος των κλειδιών που περιέχονται σε κάθε διεύθυνση και κατ' επέκταση σε κάθε φίλτρο επαφών που αποστέλλεται, εξαρτάται αποκλειστικά από το πλήθος των κόμβων στο δίκτυο. Ωστόσο, για να λειτουργεί ο μηχανισμός που παρέχει ανωνυμία, αφαιρούμε κάποια από τα έγκυρα κλειδιά που περιέχονται σε μια διεύθυνση. Αυτό αρκεί για να μπερδέψει έναν κακόβουλο χρήστη (βλέπε ενότητα 3.4). Αντί, λοιπόν, να αφαιρούμε μη έγκυρα κλειδιά, ουσιαστικά μπορούμε να αφαιρέσουμε έγκυρα κλειδιά από με κέρδος τη μείωση του μεγέθους των φίλτρων Bloom.

Η νέα παραλλαγή δημιουργεί ζητήματα σχετικά με την απόδοση του δικτύου. Στην αρχική ιδέα, όλα τα έγκυρα κλειδιά υπήρχαν στα φίλτρα διευθύνσεων των κόμβων (αφαιρούσαμε μόνο εικονικά κλειδιά). Κατά συνέπεια, ο υπολογισμός των μετρικών μπορούσε να επηρεαστεί μόνο από την ύπαρξη false positives. Στη νέα προσέγγιση αφαιρούμε ένα ποσοστό των έγκυρων κλειδιών. Αυτό αναμένουμε να μας δημιουργήσει επιπλέον λάθη στον υπολογισμό των μετρικών. Τα επιπλέον λάθη δεν θα σχετίζονται μόνο με την ύπαρξη false positives αλλά και με την ελλιπή πληροφορία που θα υπάρχει στα φίλτρα επαφών. Σε αυτή την ενότητα, λοιπόν, εξετάζουμε την απόδοση της νέας προσέγγισης συναρτήσει του ποσοστού των false positives, όταν κάθε φορά αφαιρούμε διαφορετικό πλήθος έγκυρων κλειδιών.

Στα πειράματα που εκτελέσαμε χρησιμοποιήσαμε τις ίδιες μετρικές και τις ίδιες τιμές της παραμέτρου p_{fp} (1%, 5%, 10%, 20%) για να έχουμε κοινό μέτρο σύγκρισης. Τα πειράματα έγιναν με χρήση του συνόλου επαφών Infocom05. Το ποσοστό των κλειδιών που αφαιρούμε είναι 10% και 20% αντίστοιχα.

Στο Σχήμα 4.15 απεικονίζεται το ποσοστό επιτυχούς παράδοσης πακέτων συναρτήσει της τιμής p_{fp}. Παρατηρούμε ότι, αφαιρώντας ένα ποσοστό των έγκυρων κλειδιών, το

ποσοστό επιτυχούς παράδοσης μειώνεται σε σχέση τόσο με τον SimBet όσο και με την αρχική εκδοχή του SimBet-BF. Η συμπεριφορά αυτή είναι αναμενόμενη καθώς, εκτός από τα λάθη που οφείλονται στην ύπαρξη false positives, λάθη στους υπολογισμούς προκύπτουν λόγω της ελλιπούς πληροφορίας. Η αφαίρεση κλειδιών σημαίνει ότι θα υπάρχουν πολύ περισσότερες περιπτώσεις όπου ο υπολογισμός των μετρικών δρομολόγησης θα είναι λανθασμένος. Τα λάθη αυτά αποτελούν γενίκευση της ειδικής περίπτωσης υποεκτίμησης της μετρικής similarity για την οποία μιλήσαμε στην ενότητα 3.3.3. Ωστόσο, στην περίπτωση που εξετάζουμε τώρα, είναι πολύ πιο πιθανόν να οδηγήσουν σε διαφορετικές αποφάσεις δρομολόγησης (βλέπε Πίνακας 4.8), αφού ένα σημαντικό ποσοστό κλειδιών αφαιρείται. Ωστόσο, παρά την ελλιπή πληροφορία που λαμβάνουν οι κόμβοι, το ποσοστό επιτυχούς παράδοσης πακέτων υπολείπεται λιγότερο από 2% σε σχέση με τον αρχικό αλγόριθμο, ανεξαρτήτως ποσοστού αφαίρεσης κλειδιών. Αυτό σημαίνει ότι και η λύση της αφαίρεσης έγκυρων κλειδιών από τις διευθύνσεις των κόμβων αποτελεί μια αποδεκτή λύση στο πρόβλημά μας.



Σχήμα 4.15 Ποσοστό επιτυχούς παράδοσης πακέτων συναρτήσει του ποσοστού false positive

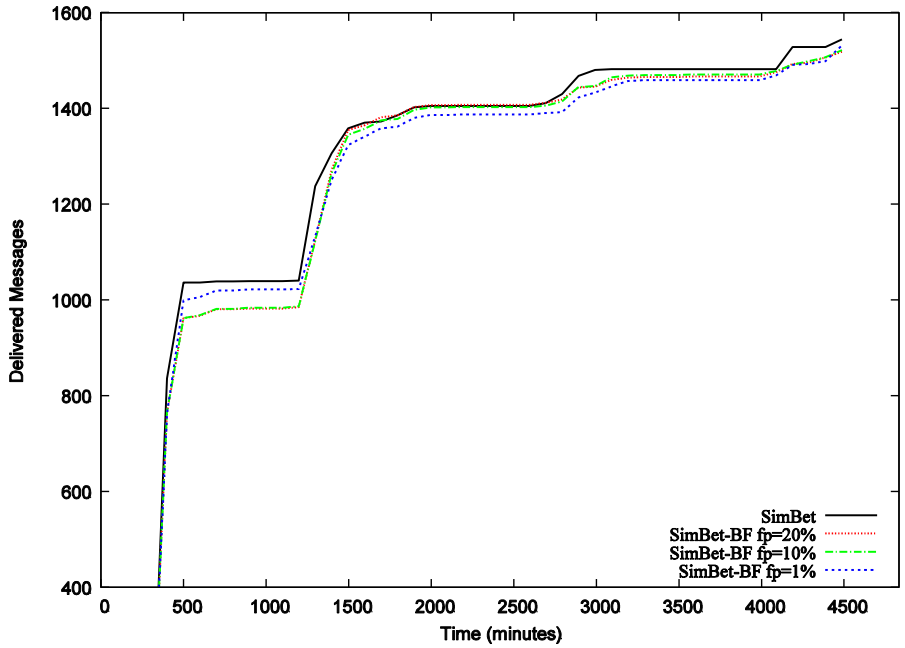
Οι διαφορετικές αποφάσεις δρομολόγησης είναι λιγότερες όταν αφαιρούμε περισσότερα κλειδιά από τα φίλτρα των διευθύνσεων. Αυτό εξηγείται αν σκεφτεί κανείς ότι αφαιρώντας περισσότερα κλειδιά μειώνεται το πραγματικό ποσοστό false

positives και έτσι αντισταθμίζονται τυχόν λάθη στους υπολογισμούς λόγω της αφαίρεσης κλειδιών.

Πίνακας 4.8 Ποσοστό διαφορετικών αποφάσεων δρομολόγησης

Fp rate (%)	Different decisions (%)	
	remove 10% of keys	remove 20% of keys
1%	19.47%	18.84%
5%	20.1%	19.26%
10%	21.29%	19.35%
20%	23.07%	19.79%

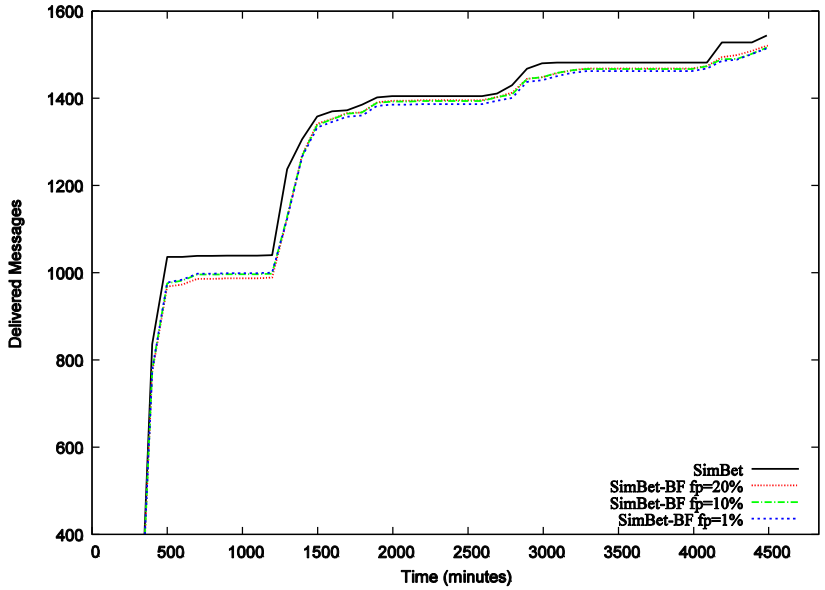
Τα λάθη στις αποφάσεις δρομολόγησης αντικατοπτρίζονται και στα επόμενα δύο διαγράμματα (Σχήματα 4.16 και 4.17) όπου φαίνεται ο αριθμός των παραδιδόμενων πακέτων σε συνάρτηση με το χρόνο.



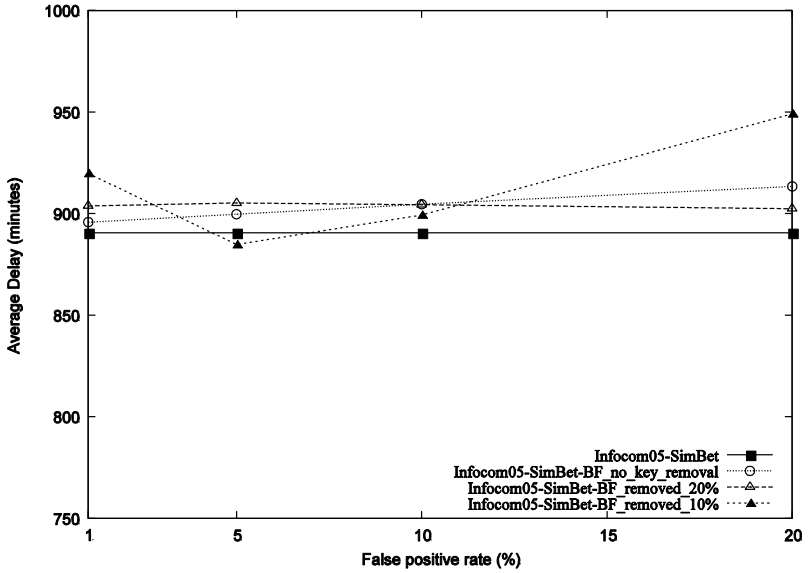
Σχήμα 4.16 Αριθμός παραδιδόμενων πακέτων συναρτήσει του χρόνου για ποσοστό αφαίρεσης κλειδιών 10% επί του συνόλου

Όπως και στα προηγούμενα πειράματα, τα παραδιδόμενα πακέτα στα ενδιάμεσα χρονικά σημεία της προσομοίωσης είναι λιγότερα από ότι στον αρχικό αλγόριθμο. Τα

λάθη στη δρομολόγηση έχουν ως αποτέλεσμα τα πακέτα να ακολουθούν διαφορετικά δρομολόγια με αποτέλεσμα να φτάνουν στο προορισμό τους με μεγαλύτερη καθυστέρηση σε σχέση με τον SimBet (βλέπε Σχήμα 4.18).



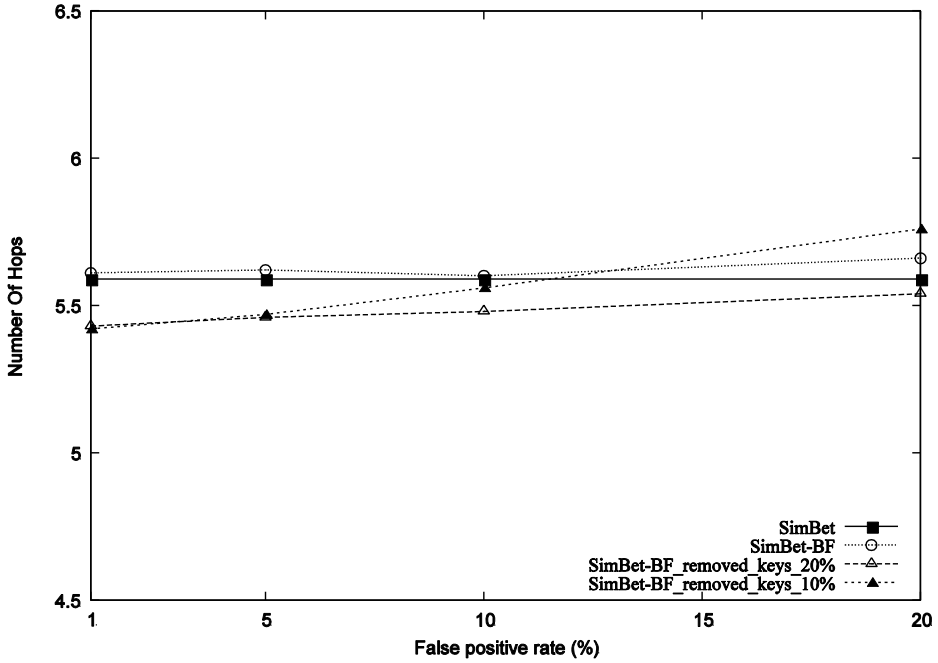
Σχήμα 4.17 Αριθμός παραδιδόμενων πακέτων συναρτήσει του χρόνου για ποσοστό αφαίρεσης κλειδιών 20% επί του συνόλου



Σχήμα 4.18 Μέση καθυστέρηση συναρτήσει του ποσοστού false positive

Η μέση καθυστέρηση, όπως φαίνεται στο Σχήμα 4.18, βλέπουμε ότι είναι αυξημένη σε σχέση με τον SimBet. Η αύξηση αυτή είναι της τάξης του 5%, δηλ. κάπως

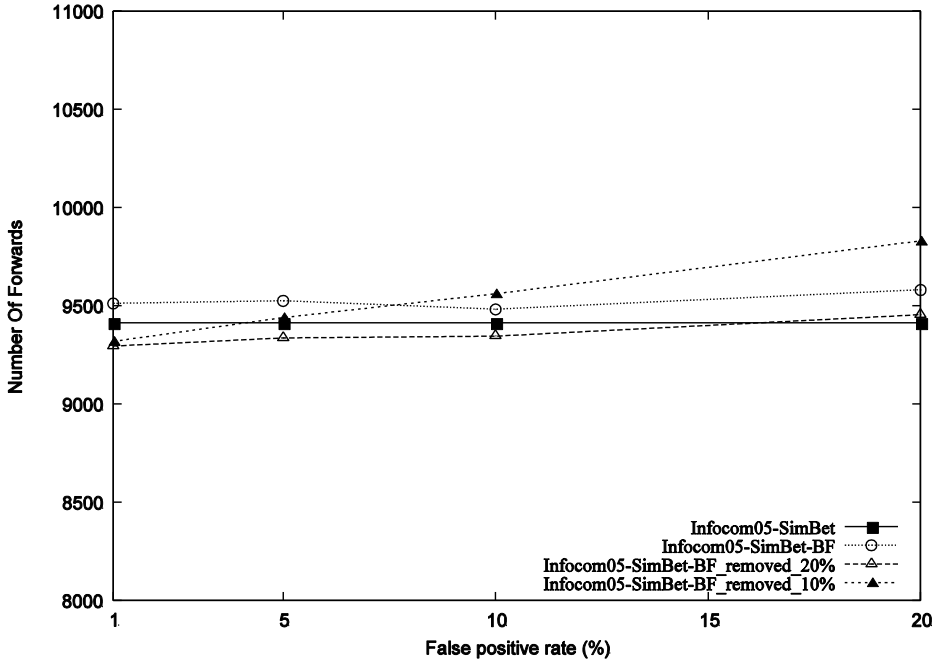
μεγαλύτερη σε σχέση με την καθυστέρηση της αρχικής έκδοσης του ανωνυμοποιημένου SimBet. Δεδομένου ότι έχουμε περισσότερα λάθη στη δρομολόγηση τα πακέτα ακολουθούν εναλλακτικές διαδρομές, διαφορετικές από τον αρχικό αλγόριθμο. Οι διαδρομές αυτές, κατά κανόνα, εισάγουν μεγαλύτερη καθυστέρηση. Επίσης, η αυξημένη καθυστέρηση μπορεί να οφείλεται και στη διατήρηση πακέτων στους κόμβους για μεγάλο χρονικό διάστημα (οι κόμβοι δεν προωθούν τα πακέτα σε άλλους κόμβους). Συνεπώς, η καθυστέρηση στην παράδοση των πακέτων είναι αυξημένη σε κάθε περίπτωση και μπορούμε να πούμε ότι είναι αναμενόμενη, χωρίς όμως, να φτάνει σε μη ανεκτά επίπεδα για το δίκτυο.



Σχήμα 4.19 Μέσος αριθμός αλμάτων συναρτήσει του ποσοστού false positive

Στο Σχήμα 4.19 φαίνεται ο μέσος αριθμός αλμάτων για τα διαφορετικά ποσοστά αφαίρεσης κλειδιών. Εδώ φαίνεται ότι όταν αφαιρούμε περισσότερα κλειδιά (20% του συνόλου) από τα φίλτρα διευθύνσεων, ο αριθμός των αλμάτων είναι σαφώς μικρότερος από τον αρχικό αλγόριθμο. Αυτός είναι ένας από τους δυο λόγους που έχουμε αύξηση στην καθυστέρηση. Σε αυτή την περίπτωση, όπως είπαμε, είναι πιο πιθανό, λόγω υποεκτίμησης της μετρικής similarity τα πακέτα να μην προωθούνται και να μένουν αποθηκευμένα στους κόμβους. Όταν αφαιρούμε λιγότερα κλειδιά (10% του συνόλου) και όσο το ποσοστό των false positives αυξάνεται, είναι πιο

πιθανόν να επιλέγουμε λάθος κόμβους για την προώθηση των πακέτων. Συνεπώς, τα πακέτα ακολουθούν δρομολόγια με περισσότερα άλματα και επομένως καθυστερεί η παράδοσή τους. Τα παραπάνω επιβεβαιώνονται και από το διάγραμμα του σχήματος 4.20 όπου εμφανίζεται ο συνολικός αριθμός προωθήσεων. Όταν αφαιρούμε περισσότερα κλειδιά (20% του συνόλου) τα λιγότερα άλματα που εκτελούν τα πακέτα συνεπάγονται και λιγότερες προωθήσεις από τους ενδιάμεσους κόμβους.



Σχήμα 4.20 Συνολικός αριθμός προωθήσεων συναρτήσει του ποσοστού false positive

Αντίθετα όταν αφαιρούμε λιγότερα κλειδιά (10% του συνόλου), τα παραπάνω άλματα σημαίνουν και περισσότερες προωθήσεις για τα πακέτα. Σε κάθε περίπτωση οι αυξήσεις σε μέσο αριθμό αλμάτων και συνολικό αριθμό προωθήσεων κυμαίνονται λίγο πάνω από 4% σε σχέση με τον αρχικό αλγόριθμο SimBet και σε καμία περίπτωση δεν αποτελούν ανασταλτικό παράγοντα εφαρμογής της μεθόδου.

ΚΕΦΑΛΑΙΟ 5. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ

5.1. Συμπεράσματα

Στα πλαίσια της παρούσας διατριβής παρουσιάσαμε μια διαφορετική προσέγγιση για την προστασία της ανωνυμίας κατά τη δικτυακή επικοινωνία. Το μοντέλο δικτύου στο οποίο αναπτύξαμε τη λύση μας είναι τα opportunistic δίκτυα. Η εξασφάλιση ανωνυμίας σε ένα τέτοιο δίκτυο είναι ένα δύσκολο ερευνητικό πρόβλημα και η έρευνα που έχει γίνει πάνω σε αυτό το ζήτημα είναι περιορισμένη. Οι μέχρι τώρα γνωστές τεχνικές χρησιμοποιούν ως βασικό εργαλείο την κρυπτογραφία, η οποία δεν ενδείκνυται ως λύση για περιβάλλοντα όπου οι κόμβοι διαθέτουν περιορισμένους υπολογιστικούς πόρους (φορητές συσκευές) και το μονοπάτι επικοινωνίας δεν είναι εκ των προτέρων γνωστό.

Στη διατριβή αυτή προτείνουμε τη χρήση φίλτρων Bloom για ανώνυμη επικοινωνία σε opportunistic δίκτυα. Σε αυτά τα δίκτυα οι υπάρχουσες τεχνικές δεν μπορούν να εφαρμοστούν επειδή το μονοπάτι επικοινωνίας δεν μπορεί να καθοριστεί εξ' αρχής από τον αποστολέα. Επίσης, με τη χρήση φίλτρων Bloom μειώνουμε το υπολογιστικό κόστος που απαιτεί η κρυπτογράφηση. Ο αλγόριθμος στον οποίο βασιστήκαμε είναι ο SimBet, ο οποίος ανήκει στους αλγορίθμους κοινωνικής δικτύωσης και η ιδέα ήταν να αντικαταστήσουμε τις παραδοσιακές δομές του με φίλτρα Bloom. Οι διευθύνσεις και οι επαφές των κόμβων που απαιτούνται για τη διαδικασία της δρομολόγησης αναπαρίστανται πλέον με φίλτρα Bloom ώστε να προφυλαχθούν από κακόβουλους

ενδιάμεσους κόμβους. Αυτός ο τρόπος αναπαράστασης μας εξασφαλίζει ότι οι κακόβουλοι κόμβοι δεν μπορούν, άμεσα ή συνδυάζοντας την πληροφορία που λαμβάνουν, να αποκαλύψουν τα άκρα της επικοινωνίας. Επιπλέον, κατά τη αλληλεπίδραση δύο κόμβων που εκτελούν το πρωτόκολλο, η επικοινωνία είναι ανώνυμη. Ένα ζήτημα που προκύπτει από τη χρήση φίλτρων Bloom είναι ότι η αντικατάσταση των πραγματικών δομών του αλγορίθμου με φίλτρα Bloom εμπεριέχει την πιθανότητα λαθών στον υπολογισμό των μετρικών betweenness και similarity. Ωστόσο, αποδείχθηκε ότι τα λάθη αυτά δεν επηρεάζουν ουσιαστικά τη λειτουργία του αλγορίθμου. Ειδικότερα, για τη μετρική betweenness αποδείχθηκε ότι είναι πρακτικά αδύνατο να γίνει λάθος στον υπολογισμό της. Για τη μετρική similarity, η οποία στον παραδοσιακό αλγόριθμο απαιτεί γνώση του προορισμού, πετυχαίνουμε υπολογισμό με μεγάλη ακρίβεια, χωρίς να έχουμε γνώση της πραγματικής διεύθυνσης του προορισμού.

Η χρήση των φίλτρων Bloom δημιουργεί και κάποια ζητήματα όσον αφορά την υλοποίησή τους σε ένα πραγματικό σύστημα. Το βασικότερο ζήτημα αφορά το μέγεθος των φίλτρων, το οποίο μπορεί να είναι αρκετά μεγάλο. Αυτό το ζήτημα μπορεί να αντιμετωπιστεί επιτυχώς υιοθετώντας εναλλακτικές προσεγγίσεις όπως τη διατήρηση των πιο πρόσφατων επαφών αντί όλων και την αποστολή των φίλτρων με μια πιο συμπαγή αναπαράσταση (αποστολή των θέσεων που έχουν τεθεί άσσοι, αντί όλου του φίλτρου). Αυτές οι εναλλακτικές φαίνεται ότι κάνουν χρηστική και υλοποιήσιμη τη λύση που προτείνουμε, χωρίς να αυξάνουν την πολυπλοκότητά της.

Στα πειράματα που διεξαγάγαμε για την αξιολόγηση της μεθόδου χρησιμοποιήσαμε πραγματικά σύνολα επαφών καταγεγραμμένα από φορητές συσκευές αντί κάποιου συνθετικού μοντέλου. Με τον τρόπο αυτό εξασφαλίσαμε πιο ρεαλιστικά αποτελέσματα. Τα αποτελέσματα δείχνουν ότι η επίδραση του ποσοστού των false positives, λόγω της χρήσης των φίλτρων Bloom, επηρεάζει ελάχιστα τη γενικότερη απόδοση του αλγορίθμου. Οι διαφορές σε σχέση με τον αρχικό αλγόριθμο SimBet για τις μετρικές που χρησιμοποιήσαμε κυμαίνονται μεταξύ 0.5% και 3% για τα περισσότερα σύνολα επαφών. Μάλιστα, σε κάποιες περιπτώσεις η λύση μας ξεπερνά οριακά την επίδοση του SimBet.

5.2. Μελλοντικές επεκτάσεις

Μια ενδιαφέρουσα επέκταση της παρούσας διατριβής θα ήταν η περαιτέρω εξέταση των ζητημάτων υλοποίησης που περιγράψαμε στην ενότητα 3.6. Συγκεκριμένα, θα είχε ενδιαφέρον η υλοποίηση του αλγορίθμου με δυνατότητα ανανέωσης της λίστας επαφών κάθε κόμβου (διαγραφή παλιών επαφών) με χρήση παραθύρου ιστορίας. Σε αυτή την περίπτωση θα χρησιμοποιούσαμε ένα σταθερό αριθμό επαφών για κάθε κόμβο (τις πιο πρόσφατες) με αποτέλεσμα το φίλτρο να κατασκευάζεται έτσι, ώστε να χωράει λιγότερες επαφές από το σύνολο των κόμβων του δικτύου και συνεπώς να έχει μικρότερο μέγεθος. Αυτό θα έκανε πιο χρηστικό τον νέο αλγόριθμο και πιθανότατα θα οδηγούσε και σε καλύτερη επίδοση του αλγορίθμου, επειδή η δρομολόγηση θα γινόταν με βάση τις πιο πρόσφατες επαφές δηλ. με χρήση ανανεωμένης (πιο πρόσφατης) πληροφορίας σχετικά με το δίκτυο.

Άλλη μια κατεύθυνση έρευνας θα ήταν η μελέτη της χρήσης των φίλτρων Bloom σε άλλες κατηγορίες πρωτοκόλλων δρομολόγησης για DTN/opportunistic δίκτυα. Ειδικότερα, πιστεύουμε ότι είναι δυνατή η χρήση φίλτρων Bloom για προστασία της ανωνυμίας στα πρωτόκολλα που στηρίζονται στον αλγόριθμο Epidemic καθώς η αλληλεπίδραση των κόμβων κατά την εκτέλεση του πρωτοκόλλου είναι παρόμοια. Αυτό που χρειάζεται είναι η προσαρμογή της αναπαράστασης των κόμβων στο εκάστοτε πρωτόκολλο ώστε ο έλεγχος των κριτηρίων δρομολόγησης να γίνεται με ακρίβεια.

ΑΝΑΦΟΡΕΣ

[1] B. Bloom. “Space/Time Trade-offs in Hash Coding with Allowable Errors”. Communications of the ACM, pp 422-426, July 1970.

[2] D. Boneh, and M. K. Franklin, “Identity-based Encryption From the Weil Pairing”, Proceedings of CRYPTO 2001, pp. 213-229, 2001.

[3] H. Choi, P. McDaniel and T. La Porta. “Privacy Preserving Communication in MANETs”. Sensor, Mesh and Ad Hoc Communications and Networks, June 2007.

[4] A Community Resource for Archiving Wireless Data at Dartmouth, http://crawdad.cs.dartmouth.edu/

[5] E. Daly and M. Haahr. “Social network analysis for routing in disconnected delay-tolerant MANETs”, Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing, pp 32-40, September 2007.

[6] C. Diot et al., Haggle project, http://www.cl.cam.ac.uk/research/srg/netos/haggle/

[7] D. Goldschlag, M. Reed and P. Syverson. “Onion Routing for Anonymous and Private Internet Connections”, Communications of the ACM. Vol. 42(2), pp 39-41, February 1999.

[8] P. Hui, J. Crowcroft and E.Yoneki. “BUBBLE Rap: social-based forwarding in delay tolerant networks”. Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing, pp 241-250, May 2008.

[9] A. Kate, G. Zaverucha and U. Hengartner, “Anonymity and Security in Delay Tolerant Networks”, Third International Conference on Security and Privacy in Communications Networks (SecureComm), September 2007.

[10] A.Keranen, J. Ott and T. Karkkainen. “The ONE Simulator for DTN Protocol Evaluation”. Proceedings of the 2nd International Conference on Simulation Tools and Techniques (SIMUTools), March 2009.

[11] A. Lindgren, A. Doria and O. Schelen. “Probabilistic routing in intermittently connected networks”. Lecture Notes in Computer Science, Vol (3126), pp 239-254, 2004.

[12] R. Lu et al, “Ecpp: Efficient Conditional Privacy-preservation Protocol for Secure Vehicular Communications”, Proceedings of INFOCOM 2008, pp. 1229-1237, April 2008.

[13] R. Lu, X. Lin and X. Shen, “SPRING: A Social-based Privacy-preserving Packet Forwarding Protocol for Vehicular Delay Tolerant Networks”, INFOCOM'10 Proceedings of the 29th conference on Information communications, pp. 632-640, 2010.

[14] Y. Onn, et. al., Privacy in the Digital Environment, pp. 1-12, Haifa Center of Law & Technology, 2005.

[15] I. Parris and T. Henderson. “Privacy-enhanced social network routing in opportunistic networks”. Pervasive Computing and Communications Workshop (PERCOM), March-May 2010.

[16] R. Ramanathan et al. “Prioritized Epidemic Routing for Opportunistic Networks”. Proceedings of the 1st international MobiSys workshop on Mobile opportunistic networking, pp 62-66, June 2007.

[17] The Reality Mining Project, http://reality.media.mit.edu/

[18] T. Spyropoulos, K. Psounis and C. S. Raghavendra. “Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks”, Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking, pp 252-259, August 2005.

[19] A. Vahdat, D. Becker, “Epidemic Routing for Partially-Connected Ad Hoc Networks”, Duke Tech Report CS-2000-06, 2000.

[20] Y. Zhu and Y. Hu. “Making Peer-to-Peer Anonymous Routing Resilient to Failures”. IEEE Parallel and Distributed Processing Symposium (IPDPS), March 2007.

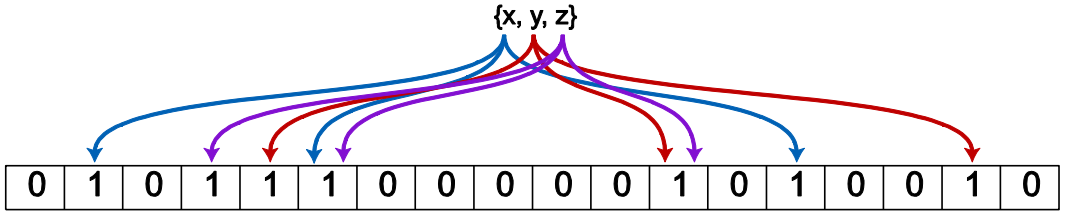
ΠΑΡΑΡΤΗΜΑ

Σε αυτό το παράρτημα περιγράφουμε τη δομή και τον τρόπο κατασκευής των φίλτρων Bloom, στα οποία βασίστηκε η παρούσα διατριβή.

Φίλτρα Bloom

Τα φίλτρα Bloom είναι μια ειδική δομή δεδομένων η οποία χρησιμεύει στην αναπαράσταση των στοιχείων ενός συνόλου S με χρήση ενός μονοδιάστατου δυαδικού πίνακα. Επίσης, μας παρέχει τη δυνατότητα να ελέγξουμε αν ένα στοιχείο x περιέχεται στο φίλτρο, χωρίς να έχουμε πλήρη εικόνα των στοιχείων του φίλτρου. Ένα φίλτρο Bloom είναι ένας δυαδικός μονοδιάστατος πίνακας M μεγέθους m, του οποίου, αρχικά, κάθε θέση έχει τιμή 0. Για να εισάγουμε στοιχεία ενός συνόλου στο φίλτρο, πρέπει να ορίσουμε *k* συναρτήσεις κατακερματισμού (hash functions) $h_i(x)$, i=1,…,k.

Η εισαγωγή ενός στοιχείου στο φίλτρο γίνεται ως εξής: έστω ένα στοιχείο x. Υπολογίζουμε για καθεμία από τις k συναρτήσεις κατακερματισμού το $pos_i=h_i(x)$. Στην αντίστοιχη θέση του μονοδιάστατου πίνακα M θέτουμε την τιμή 1 δηλ. $M[pos_i]=1$, i=1,…,k. Έτσι, το στοιχείο x αναπαρίσταται στον πίνακα M με k άσους.



Σχήμα Π.1 Φίλτρο Bloom με 3 στοιχεία

Για να ελέγξουμε αν ένα στοιχείο w περιέχεται στο φίλτρο Bloom (membership check), τροφοδοτούμε καθεμία από τις συναρτήσεις κατακερματισμού με το w και για κάθε ακέραιο pos_j που επιστρέφει η κάθε συνάρτηση, ελέγχουμε αν η θέση $M[pos_j]$ είναι ίση με 1. Αν έστω και για ένα j έχουμε $M[pos_j]=0$, τότε το w δεν περιέχεται στο φίλτρο. Αν για όλα τα $M[pos_j]$ ισχύει $M[pos_j]=1$ τότε είτε το στοιχείο w περιέχεται στο φίλτρο, είτε οι αντίστοιχες θέσεις του M τέθηκαν ίσες με 1 κατά την εισαγωγή άλλων στοιχείων. Άρα, το w περιέχεται στο φίλτρο με κάποια πιθανότητα.

Από τα παραπάνω συμπεραίνουμε ότι χρησιμοποιώντας φίλτρα Bloom υπάρχει πιθανότητα να έχουμε *false positives*, δηλ. να λαμβάνουμε εσφαλμένη θετική απάντηση για την ύπαρξη ενός στοιχείου. Από την άλλη όμως, αποκλείεται η πιθανότητα να έχουμε *false negatives*, δηλ. να λάβουμε εσφαλμένη αρνητική απάντηση. Αυτή η ιδιότητα των φίλτρων Bloom προκύπτει από το γεγονός ότι οι συναρτήσεις κατακερματισμού δεν μπορούν να είναι ιδανικές, δηλ. να μας δίνουν πάντα διαφορετικές θέσεις (ακεραίους) για κάθε στοιχείο που κατακερματίζεται. Μπορούμε, όμως, να επιλέξουμε το μέγεθος του φίλτρου και τον αριθμό των συναρτήσεων κατακερματισμού με τέτοιο τρόπο, ώστε η πιθανότητα να έχουμε false positives να είναι πολύ μικρή. Συγκεκριμένα, αν m είναι το μέγεθος του φίλτρου και n τα στοιχεία που κατακερματίζονται με χρήση k συναρτήσεων, τότε η πιθανότητα εμφάνισης false positive, p, αποδεικνύεται ότι είναι ίση με

$$p = \left(1 - \left[1 - \frac{1}{m}\right]^{kn}\right)^k \approx \left(1 - e^{-kn/m}\right)^k \qquad \text{Εξ. Π.1}$$

Αν θεωρήσουμε ένα όριο για τα false positives, έστω p, τότε μπορούμε να υπολογίσουμε το μέγεθος m του φίλτρου με βάση τον παρακάτω τύπο

$$m = -\frac{n \ln p}{(\ln 2)^2} \qquad \text{Εξ. Π.2}$$

όπου n το πλήθος του των στοιχείων στο φίλτρο.

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Ο Βασίλειος Μπούργος γεννήθηκε το 1985 στην Άρτα. Αποφοίτησε από το Ενιαίο Λύκειο Άνω Καλεντίνης το 2003. Το 2004 εισήχθη στο προπτυχιακό πρόγραμμα σπουδών του Τμήματος Πληροφορικής του Πανεπιστημίου Ιωαννίνων. Το Μάρτιο του 2009 ολοκλήρωσε τις σπουδές του και στη συνέχεια παρακολούθησε το μεταπτυχιακό πρόγραμμα σπουδών του ίδιου τμήματος. Τα ερευνητικά του ενδιαφέροντα εστιάζονται στους τομείς των δικτύων υπολογιστών και συγκεκριμένα στην εξασφάλιση ιδιωτικότητας σε δίκτυα κινητών κόμβων.