ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΔΙΔΑΚΤΙΚΗ ΤΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΨΗΦΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

ΚΑΤΕΥΘΥΝΣΗ:ΨΗΦΙΑΚΕΣ ΕΠΙΚΟΙΝΩΝΙΕΣ ΚΑΙ ΔΙΚΤΥΑ

ΜΟΝΤΕΛΑ ΕΞΑΠΛΩΣΗΣ ΣΚΟΥΛΗΚΙΩΝ

ΘΕΟΔΩΡΟΠΟΥΛΟΣ ΑΝΔΡΕΑΣ

Αθήνα, Ιούνιος 2010

Αφιερώνεται
στην αδελφή μου Βασιλική

0.0 Περίληψη

Τα σκουλήκια,στις μέρες μας, αποτελούν μία από τις σοβαρότερες απειλές που στοχεύουν σε συσκευές που επικοινωνούν μέσω δικτύου.Η ανάλυση και μοντελοποίηση της εξάπλωσης του συγκεκριμένου είδους κακόβουλου λογισμικού,συντελεί στην εύρεση μεθόδων άμυνας των παραπάνω συσκευών,μεθόδων περιορισμού της μόλυνσης και τελικά στην προστασία του χρήστη.Στα πλαίσια της συγκεκριμένης διπλωματικής εργασίας,θα αναλυθούν τα τοπολογικά χαρακτηριστικά των πολύπλοκων δικτύων μέσω των οποίων θα φανεί η επιρροή που ασκούν στην διαδικασία εξάπλωσης των σκουληκιών σε πραγματικά δίκτυα,θα γίνει διαχωρισμός των σκουληκιών με βάση τα κυριότερα χαρακτηριστικά τους,θα αναλυθούν και θα αξιολογηθούν τα επιδημικά και άλλα ήδη γνωστά μοντέλα εξάπλωσης σκουληκιών καθώς επίσης θα παρουσιαστούν μοντέλα εξάπλωσης σκουληκιών σε γνωστά πραγματικά δίκτυα και θα διαπιστωθεί η ακρίβεια τους βάση προσομοιώσεων.Τελικά,θα προταθεί ένα μοντέλο εξάπλωσης XSS σκουληκιών σε online κοινωνικά δίκτυα που βασίζεται στο μοντέλο εξάπλωσης two-factor και θα γίνει προσομοίωση αυτού σε Matlab.

Λέξεις-κλειδιά:πολύπλοκα δίκτυα,σκουλήκια,μοντέλα εξάπλωσης,
επιδημιολογικά μοντέλα.

0.1 Abstract

Nowadays, the worm is one of the major threats of devices that communicate through network.The analysis and modeling of the spread of such kind of malware, conduces to the finding of defense methods, infection containment methods and as a result, user protection.In this Master's thesis, the topological characteristics of complex networks will be analyzed, through which will appear the amount of influence they have on the spreading procedure of worms in real networks.In addition, there will be a separation of worms based on their characteristics, an analysis and evaluation of epidemic and other already known worm propagation models, as well as, there will be a presentation of worm propagation models in several real networks which will be evaluated.Finally,we will propose an XSS worm propagation model in online social networks which is based on the two-factor worm model and it will be simulated on Matlab.

Keywords:complex networks,worms,propagation models,
epidemic models.

Ευχαριστίες:

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον Λέκτορα κ. Χρήστο Ξενάκη για την εμπιστοσύνη που μου έδειξε κατά την ανάθεση της συγκεκριμένης διπλωματικής εργασίας καθώς επίσης και για την επίβλεψη και τη βοήθεια που μου παρείχε για την ολοκλήρωση αυτής. Θα ήθελα επιπλέον να ευχαριστήσω τον Δρ.Χριστόφορο Νταντογιάν για την εξαιρετική καθοδήγηση και τις πολύτιμες υποδείξεις του καθ'όλη τη διάρκεια εκπόνησης της.

Τέλος, θα ήθελα να εκφράσω την ευγνωμοσύνη μου στην οικογένεια μου για την υποστήριξη και την υπομονή τους καθ'όλη τη διάρκεια των σπουδών μου.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

1 Θεωρία πολύπλοκων δικτύων (complex network theory):

1.1 Εισαγωγικά :

Αυτή τη στιγμή τα πολύπλοκα δίκτυα μελετώνται σε πολλούς τομείς της επιστήμης και αυτό οφείλεται στο γεγονός ότι πολλά συστήματα στον πραγματικό κόσμο μπορούν να περιγραφούν με αρκετά μεγάλη ακρίβεια από αυτού του είδους τα δίκτυα.Μερικά παραδείγματα τέτοιων συστημάτων είναι το Internet,το World Wide Web(WWW)),ο ανθρώπινος εγκέφαλος και η παγκόσμια οικονομία.Τα πολύπλοκα δίκτυα είναι δομές που αποτελούνται από κόμβους(που παριστάνουν τις οντότητες του συστήματος) οι οποίοι συνδέονται μεταξύ τους με ακμές(που παριστάνουν τις σχέσεις αλληλεξάρτησης μεταξύ των οντοτήτων).Η παρουσία των δικτύων αυτών σε όλους τους τομείς της επιστήμης και της τεχνολογίας οδήγησε στην εμφάνιση ερευνητικών προβληματισμών που σχετίζονται με τον τρόπο με τον οποίο η δομή του δικτύου επιρρεάζει τις δυναμικές συμπεριφορές του.

Για πάνω από έναν αιώνα η μοντελοποίηση φυσικών και μη συστημάτων και διαδικασιών γινόταν κάτω από την υπόθεση ότι οι σχέσεις αλληλεξάρτησης,ανάμεσα στις οντότητες ενός συστήματος ή μίας διαδικασίας,μπορούν να αναπαρασταθούν από μία απλή δομή όπως για παράδειγμα τα Ευκλείδια πλέγματα(Euclidian lattices).

Στα τέλη του 1950,δύο μαθηματικοί οι Erdos και Renyi (ER),περιέγραψαν ένα δίκτυο με πολύπλοκη τοπολογία χρησιμοποιώντας έναν τυχαίο γράφο.Παρόλο που διαισθητικά ήταν εμφανές ότι τα πολύπλοκα δίκτυα του πραγματικού κόσμου δεν ήταν ούτε εντελώς τακτικά(regular) ούτε εντελώς τυχαία,το μοντέλο ER τυχαίου γράφου μονοπώλησε το ενδιαφέρον της επιστημονικής κοινότητας για περίπου μισό αιώνα,κυρίως λόγω έλλειψης μεγάλης υπολογιστικής ισχύος και λεπτομερών πληροφοριών για την τοπολογία των δικτύων μεγάλης κλίμακας του πραγματικού κόσμου.

Το 1998,οι Watts και Strogatz(WS) εισήγαγαν την ιδέα του small-world δικτύου με σκοπό να περιγράψουν τη μετάβαση από ένα τακτικό πλέγμα(regular lattice) σε τυχαίο γράφο(random graph).Αξίζει να σημειωθεί ότι το small-world φαινόμενο είναι γενικά πολύ συνηθισμένο.Ένα παράδειγμα από την καθημερινότητα

είναι όταν συναντάμε κάποιον ξένο και ύστερα από καποιά ώρα διαπιστώνουμε ότι έχουμε έναν τουλάχιστον κοινό γνωστό.Το φαινόμενο αυτό,αποτελεί κύριο χαρακτηριστικό πολλών δικτύων στον πραγματικό κόσμο.

Κοινό χαρακτηριστικό του ER μοντέλου και του WS μοντέλου είναι ότι κάθε κόμβος του δικτύου έχει περίπου τον ίδιο αριθμό συνδέσεων γι'αυτό ονομάζονται και ομογενή ή εκθετικά δίκτυα.

Μία ακόμα πρόσφατη,σημαντική ανακάλυψη είναι ότι παρατηρήθηκε ότι πολλά πολύπλοκα δίκτυα μεγάλης κλίμακας είναι scale-free.Με τον όρο scale-free εννοούμε ότι η συνδετικότητα τους ακολουθεί power law κατανομή που είναι ανεξάρτητη από την κλίμακα του δικτύου.Βασικό χαρακτηριστικό ενός scale-free δικτύου είναι η ανομοιογένειά του.Σε αντίθεση με τα μοντέλα δικτύων ER και WS,στα scale-free οι περισσότεροι κόμβοι έχουν πολύ λίγες συνδέσεις και λίγοι κόμβοι έχουν πολλές συνδέσεις.

1.2 Τρεις βασικές ιδέες για την ανάπτυξη πολύπλοκων δικτύων:

Τα τρία στοιχεία που παίζουν καθοριστικότατο ρόλο στην ανάπτυξη της θεωρίας πολύπλοκων δικτύων είναι:

α) το μέσο μήκος μονοπατιού

β) ο συντελεστής ομαδοποίησης

γ) η κατανομή βαθμού

Οι Watts και Strogatz ήθελαν να κατασκευάσουν ένα μοντέλο δικτύου με μικρό μέσο μήκος μονοπατιού(όπως στους τυχαίους γράφους) και σχετικά μεγάλο συντελεστή ομαδοποίησης(όπως στα τακτικά πλέγματα) και έτσι κατέληξαν στο σημερινό μοντέλο δικτύου(small-world network).

α) Μέσο Μήκος Μονοπατιού L

Το μέσο μήκος μονοπατιού L ενός δικτύου ορίζεται σαν ο μέσος όρος των αποστάσεων μεταξύ όλων των ζευγαριών του δικτύου και καθορίζει το ουσιαστικό μέγεθος του δικτύου.Αξιοσημείωτο είναι ότι το μέσο μήκος μονοπατιού στα περισσότερα πραγματικά πολύπλοκα δίκτυα είναι σχετικά μικρό.

β)Συντελεστής Ομαδοποίησης C

Ο συντελεστής ομαδοποίησης C ενός δικτύου ορίζεται σαν ο μέσος όρος των συντελεστών ομαδοποίησης C_i όλων των κόμβων ενός δικτύου,όπου C_i είναι ο λόγος του αριθμού των ακμών που ενώνουν τους γειτονικούς κόμβους ενός κόμβου i προς τον αριθμό των ακμών που ενώνουν τον κόμβο i με τους γειτονικούς του.

Είναι προφανές ότι σε όλα τα δίκτυα ισχύει C<=1.Στην ειδική περίπτωση που C=1 τότε το δίκτυο είναι πλήρως συνδεδεμένο(globally coupled).Σε ένα τελείως τυχαίο δίκτυο με N κόμβους έχουμε C~1/N που σημαίνει ότι όταν το N είναι πολύ μεγάλο το C γίνεται πολύ μικρό το οποίο έρχεται σε αντίθεση με τα πραγματικά δίκτυα,αφού τα περισσότερα έχουν μία τάση προς την ομαδοποίηση.Παρόλα αυτά,είναι πολύ μικρότερο του 1(των πλεγμάτων)που σημαίνει ότι τα πραγματικά πολύπλοκα δίκτυα δεν πρέπει να αντιμετωπίζονται ούτε σαν τυχαίοι γράφοι,ούτε σαν πλήρως συνδεδεμένα πλέγματα.

γ)Κατανομή Βαθμού

Ίσως το πιο σημαντικό χαρακτηριστικό ενός κόμβου είναι ο βαθμός του.Ο βαθμός k_i ενός κόμβου i ορίζεται σαν ο συνολικός αριθμός των συνδέσεων του και ο μέσος όρος των βαθμών όλων των κόμβων του δικτύου,μας δίνει το μέσο βαθμό όλου του δικτύου που συμβολίζεται με <k>.Η κατανομή των βαθμών των κόμβων στο δίκτυο γίνεται βάσει της συνάρτησης κατανομής P(k) όπου εκφράζει την πιθανότητα ένας τυχαία επιλεγμένος κόμβος να έχει ακριβώς k ακμές.

Τα τελευταία χρόνια πολλά εμπειρικά αποτελέσματα έδειξαν ότι για τα περισσότερα πραγματικά δίκτυα μεγάλης κλίμακας η κατανομή βαθμού των κόμβων απέχει σημαντικά από την κατανομή Poisson(την οποία ακολουθεί η κατανομή βαθμού σε τυχαία δίκτυα).Συγκεκριμένα,για έναν αριθμό δικτύων έχει αποδειχθεί ότι η κατανομή βαθμού ακολουθεί power-law κατανομή η οποία μειώνεται πιο σταδιακά από μία εκθετική.Λόγω του γεγονότος ότι η συγκεκριμένη κατανομή είναι ανεξάρτητη από την κλίμακα του δικτύου,γι'αυτό ένα δίκτυο με την παραπάνω κατανομή βαθμού ονομάζεται scale-free δίκτυο.

Παρακάτω θα αναφερθούμε στα τρια στοιχεία που προαναφέρθηκαν,στα πλαίσια των κυριότερων κατηγοριών δικτύων,και θα σχολιαστεί το κατά πόσο είναι ακριβής η αναπαράσταση των πραγματικών δικτύων κάνοντας χρήση αυτών των

μοντέλων.

1.2.1 Τακτικά συνδεδεμένα (regular coupled) δίκτυα:

Διαισθητικά ένα πλήρως συνδεδεμένο δίκτυο έχει το μικρότερο μέσο μήκος μονοπατιού και τον μεγαλύτερο συντελεστή ομαδοποίησης,χαρακτηριστικά που υπάρχουν σε πολλά πραγματικά δίκτυα.Παρόλα αυτά,είναι εύκολο να εντοπιστούν οι περιορισμοί του συγκεκριμένου μοντέλου δικτύου.Συγκεκριμένα,αν θεωρήσουμε ένα πλήρως συνδεδεμένο δίκτυο με N κόμβους τότε υπάρχουν N(N-1)/2 ακμές που συνδέουν τους κόμβους.Το γεγονός αύτο έρχεται σε αντίθεση με την δομή των πραγματικών δικτύων όπου τα περισσότερα από αυτά έχουν ακμές,το πλήθος των οποίων είναι N τάξης και όχι N^2 .Δηλαδή είναι πιο αραιά από ένα πλήρως συνδεδεμένο δίκτυο.

΄Ενα ευρέως μελετημένο αραιό,τακτικό μοντέλο δικτύου είναι το συνδεδεμένο δίκτυο κοντινότερου γείτονα(nearest neighbour coupled).Πρόκειται για τακτικό γράφο,όπου κάθε κόμβος του συνδέεται με μερικούς από τους γείτονες του (και όχι με όλους).

Υποθέτοντας ότι το δίκτυο κοντινότερου γείτονα αποτελείται από N κόμβους τότε αυτοί είναι τοποθετημένοι έτσι ώστε να σχηματίζουν δακτύλιο.Στον δακτύλιο αυτό,κάθε κόμβος i του δικτύου είναι συνδεδεμένος με τους γείτονες του i=1,2,....,K/2 όπου K είναι άρτιος.Για μεγάλο τέτοιο K το δίκτυο είναι πολύ ομαδοποιημένο και ο συντελεστής ομαδοποίησης του είναι περίπου C=3/4.Όμως ένα τέτοιο δίκτυο δεν είναι small-world.Αντιθέτως το μέσο μήκος μονοπατιού L γίνεται αρκετά μεγάλο όσο N->οο και συνεπώς η περιγραφή ενός πραγματικού δικτύου χρησιμοποιώντας το συγκεκριμένο μοντέλο δεν ενδείκνειται.Ίσως το μοναδικό μοντέλο δικτύου που ανήκει στα τακτικά συνδεδεμένα δίκτυα,είναι αραιό,ομαδοποιημένο,έχει μικρό μέσο μήκος μονοπατιού και μπορεί να περιγράψει ένα πραγματικό δίκτυο μεγάλης κλίμακας είναι το star-shaped δίκτυο που στην πραγματικότητα ούτε αυτό είναι ιδανικό καθώς το σχήμα του απέχει πολύ από των πραγματικών δικτύων.

1.2.2 Τυχαίοι γράφοι (Random Graphs):

Στο αντίθετο άκρο από αυτό που βρίσκονται τα εντελώς τακτικά δίκτυα,βρίσκονται οι εντελώς τυχαίοι γράφοι οι οποίοι μελετήθηκαν από τους μαθηματικούς Erdos και Renyi[2].Στο μοντέλο αυτό οι ER θεώρησαν ότι έχουμε N κόμβους τυχαία κατανεμημένους στο χώρο και με ίση πιθανότητα p συνδέεται κάθε ζεύγος αυτών με μία ακμή άρα συνολικά θα υπάρχουν pN(n-1)/2 ακμές.Αν οι κόμβοι παρομοιαστούν με κουμπιά,οι ER απέδειξαν ότι αν η πιθανότητα p είναι μεγαλύτερη από μία τιμή κατωφλίου ~lnN/N τότε σχεδόν όλοι οι κόμβοι είναι συνδεδεμένοι,που στην περίπτωση των κουμπιών σημαίνει ότι αν σηκωθεί ένα κουμπί μετά από τυχαία επιλογή τότε θα σηκωθούν όλα τα κουμπιά.

Ο μέσος βαθμός ενός τυχαίου γράφου είναι <k>=p(N-1)~=pN.Αν L_{rand} είναι το μέσο μήκος μονοπατιού τότε διασθητικά περίπου <k> ^{Lrand} κόμβοι θα απέχουν απόσταση L_{rand} ή μικρότερη άρα N~<k> ^{Lrand} και συνεπώς L_{rand} ~lnN/ln<k>.Από τη συγκεκριμένη σχέση γίνεται σαφές ότι η λογαριθμική αύξηση στο L_{rand} με το μέγεθος του δικτύου N αποτελεί ένα συνηθισμένο small-world φαινόμενο γιατί το lnN αυξάνει αργά σε σχέση με το N και έτσι το L_{rand} παραμένει αρκετά μικρό ακόμα και σε μεγάλα δίκτυα.

Σχετικά με τον συντελεστή ομαδοποίησης C του μοντέλου ER αποδείχθηκε ότι είναι C=p=<k>/N<<1,που σημαίνει ότι γενικά δεν εμφανίζεται ομαδοποίηση. Στην πραγματικότητα για μεγάλο N,ο αλγόριθμος ER δημιουργεί ένα ομογενές δίκτυο του οποίου η συνδετικότητα ακολουθεί κατα προσέγγιση κατανομή Poisson. Συνοψίζοντας, ενώ το ER μοντέλο έχει small-world χαρακτηριστικά, δεν είναι ομαδοποιημένο.

1.2.3 Small-World μοντέλα:

Βάσει των προηγουμενων το ER μοντέλο και το μοντέλο τακτικού πλέγματος αποτυγχάνουν να αναπαραστήσουν μερικά σημαντικά χαρακτηριστικά πολλών πραγματικών δικτύων.Εξάλλου τα περισσότερα πραγματικά δίκτυα δεν είναι ούτε εντελώς τυχαία,ούτε εντελώς τακτικά.Οι Watts και Strogatz (WS) εισήγαγαν το

μοντέλο small-world [3] ώστε να περιγράψουν τη μετάβαση από το τακτικό πλέγμα στον τυχαίο γράφο.Αρχικά πήραν ένα συνδεδεμένο δίκτυο κοντινότερου γείτονα αποτελούμενο από N κόμβους τοποθετημένους σε σχήμα δακτυλίου.Τυχαία επανένωσαν κάθε ακμή του δικτύου με πιθανότητα p,την οποία διαφοροποιούσαν με τέτοιο τρόπο ώστε να φαίνεται η μετάβαση από την τάξη(p=0) στην τυχαιότητα(p=1).Κατά τη διαδικασία επανένωσης το ένα άκρο μίας ακμής μεταφέρεται σε έναν κόμβο που επιλέγεται τυχαία από το δίκτυο,με τον περιορισμό ότι οποιοιδήποτε δύο διαφορετικοί κόμβοι δεν μπορούν να έχουν παραπάνω από μία σύνδεση μεταξύ τους και ότι κάθε κόμβος δεν μπορεί να συνδέεται με τον εαυτό του.Με την διαδικασία αυτή εισάγονται pNk/2 μεγάλης εμβέλειας ακμές,οι οποίες συνδέουν κόμβους,που υπό άλλες συνθήκες θα ανήκαν σε διαφορετικές γειτονιές.

Οι συμπεριφορές του συντελεστη ομαδοποίησης C και του μέσου μήκους μονοπατιού L στο WS small-world μοντέλο μπορούν να θεωρηθούν συνάρτηση της πιθανότητας επανένωσης p (δηλ. C(p),L(p)).Ένα τακτικό πλέγμα δακτυλίου (p=0) είναι ομαδοποιημένο αφού C(0)=3/4 αλλά έχει μεγάλο μέσο μήκος μονοπατιού αφού L(0)=N/2k >>1 .Στο παραπάνω πλέγμα αν γίνουν αρκετές τυχαίες επανενώσεις το μέσο μήκος μονοπατιού θα μειωθεί σημαντικά από την μία και από την άλλη η τοπική ομαδοποίηση του δικτύου δεν θα αλλάξει ,καταλήγοντας στο WS small-world μοντέλο.

Το WS small-world μοντέλο μπορεί να θεωρηθεί σαν ομογενές δίκτυο (όπως και το ER μοντέλο) στο οποίο όλοι οι κόμβοι έχουν περίπου τον ίδιο αριθμό συνδέσεων.Μία παραλλαγή του WS μοντέλου προτάθηκε από τους Newman-Watts[4](NW small-world μοντέλο)σύμφωνα με την οποία αντί καποιος να διακόψει τη σύνδεση μεταξύ δύο γειτονικών κόμβων ώστε να γίνει η διαδικασία της επανένωσης(WS μοντέλο),απλά την αφήνει ως έχει και προσθέτει καινούριες συνδέσεις μεταξύ δύο κόμβων με πιθανότητα p.Το NW μοντέλο εκφυλίζεται σε συνδεδεμένο δίκτυο κοντινότερου γείτονα για p=0,γίνεται ισοδύναμο με ένα πλήρως συνδεδεμένο δίκτυο για p=1 και για επαρκώς μικρό p και επαρκώς μεγάλο N γίνεται ισοδύναμο με το WS.Σήμερα και τα δύο προαναφερθέντα μοντέλα ονομάζονται απλώς small-world μοντέλα.

1.2.4 Scale-free μοντέλα:

Κοινό χαρακτηριστικό του ES και του WS μοντέλου,όπως αναφέρθηκε και προηγουμένως,είναι ότι η κατανομή συνδετικότητας του δικτύου είναι ομογενής δηλαδή παίρνει ένα μέγιστο σε μία μέση τιμή και ύστερα μειώνεται εκθετικά.Τα συγκεκριμένα δίκτυα ονομάζονται και εκθετικά.

Πρόσφατη ανακάλυψη στον τομέα των πολύπλοκων δικτύων αποτελεί η διαπίστωση ότι ένας αριθμός πολύπλοκων δικτύων μεγάλης κλίμακας συμπεριλαμβανομένου του Internet και του World Wide Web είναι scale-free και οι κατανομές συνδετικότητας τους έχουν power-law μορφη.

Για να γίνει κατανοητή η προέλευση της power-law κατανομής βαθμού,οι Barabasi και Albert (BA) πρότειναν ένα διαφορετικό δικτυακό μοντέλο[5,6].Σύμφωνα με τους Barabasi και Albert τα προηγούμενα μοντέλα δεν λάμβαναν υπόψιν τους δύο πολύ σημαντικά χαρακτηριστικά των πολύπλοκων δικτύων.Πρώτον,ότι τα πραγματικά δίκτυα είναι ανοιχτά και σχηματίζονται δυναμικά με συνεχή προσθήκη νέων κόμβων.Παρόλα αυτά,τα ήδη προτεινόμενα μοντέλα είναι στατικά,με την έννοια ότι ενώ νέες ακμές προστίθενται και επανατακτοποιούνται,ο αριθμός των κόμβων είναι σταθερός.Τέλος,οι τυχαίοι γράφοι και τα μοντέλα small-world υποθέτουν ομοιόμορφες πιθανότητες όταν δημιουργούν νέες ακμές,κάτι το οποίο δεν είναι ρεαλιστικό.Για παράδειγμα,αν θεωρήσουμε γνωστες ιστοσελίδες που έχουν ήδη πολλές συνδέσεις,τότε είναι πιθανότατο να αποκτήσουν ακόμη περισσότερες.Αυτό το φαινόμενο είναι γνωστό και σαν "ο πλούσιος γίνεται πλουσιότερος" και δεν λαμβάνεται υπόψιν στα προηγούμενα μοντέλα.

Η πρόταση του μοντέλου των BA βασίστηκε στα γεγονότα ότι από τη στιγμή που καινούριοι κόμβοι προστίθενται στο δίκτυο,αυτό αυξάνεται και ότι καινούριοι κόμβοι συνδέονται σε ήδη υπάρχοντες κόμβους με μεγάλο αριθμό συνδέσεων.

Η ονομασία scale-free του συγκεκριμένου μοντέλου προέρχεται από το γεγονός ότι όσο μεγαλώνει το δίκτυο (δηλαδή μεγαλώνει η κλίμακά του),η κατανομή βαθμού του δεν αλλάζει.Συνεχίζει να περιγράφεται από power-law με εκθέτη -3 και συνεπώς η πιθανότητα να βρεις έναν κόμβο με k ακμές είναι ανάλογη του k^{-3}.

Αριθμητικά αποτελέσματα έδειξαν ότι σε σύγκριση με έναν τυχαίο γράφο με το ίδιο μέγεθος και τον ίδιο μέσο βαθμό,το μέσο μήκος μονοπατιού του scale-free

μοντέλου είναι μικρότερο και ο συντελεστής ομαδοποίησης πολύ μεγαλύτερος.Από αυτό γίνεται σαφές ότι η ύπαρξη κάποιων "μεγάλων" κόμβων με μεγάλο αριθμό συνδέσεων παίζει σημαντικό ρόλο στο να έρθουν οι άλλοι κόμβοι του δικτύου κοντά ο ένας στον άλλο.Παρόλα αυτά,δεν υπάρχει αναλυτική μέθοδος πρόβλεψης του μέσου μήκους μονοπατιού και του συντελεστή ομαδοποίησης σε scale-free μοντέλο.

1.3 Ευαισθησία πολύπλοκων δικτύων στις επιθέσεις:

Παραθέτωντας το παρακάτω παράδειγμα γίνεται σαφές το κύριο μειονέκτημα των πολύπλοκων δικτύων.Έστω ένα πλήρως συνδεδεμένο,μεγάλης κλίμακας δίκτυο.Αν αφαιρεθεί ένας κόμβος από το δίκτυο αυτό,συνεπάγεται και η ρίξη όλων των συνδέσεων του.Στην περίπτωση που υπάρχουν περισσότερα από ένα μονοπάτια που συνδέουν δύο κόμβους i και j , η ρίξη ενός εξ αυτών σημαίνει αυτόματα την αύξηση της μεταξύ τους απόστασης d_{ij} και κατ'επέκταση την αύξηση του μέσου μήκους μονοπατιού όλου του δικτύου.Σε πιο σοβαρές περιπτώσεις,που υπάρχει μόνο ένα μονοπάτι ανάμεσα στους δύο κόμβους,τότε η ρίξη του συγκεκριμένου μονοπατιού σημαίνει και την αποσύνδεση των κόμβων μεταξύ τους.

Η συνδετικότητα ενός δικτύου είναι ανεκτική στα λάθη αν περιέχει μία πολύ μεγάλη ομάδα στην οποία περιλαμβάνονται πολλοί κόμβοι,ακόμα και μετά την αφαίρεση ενός ποσοστού αυτών.

Στο Internet βρέθηκε ότι αν παραπάνω από το 80% των κόμβων παραλειφθεί τότε είναι πιθανό το δίκτυο να μην καταρρεύσει.Παρόλα αυτά,αν γίνει επίθεση σε κάποιους κόμβους κλειδιά δηλαδή σε κόμβους με πάρα πολλές συνδέσεις τότε το παραπάνω αποτέλεσμα μπορεί επέλθει παραλείποντας ένα πολύ μικρότερο ποσοστό κόμβων.

Έχει αποδειχθεί ότι τέτοια ανεκτικότητα σε σφάλματα και ευαισθησία στις επιθέσεις είναι κύριο χαρακτηριστικό των scale-free δικτύων και προέρχονται από την ανομοιογένειά των βαθμών σε αυτά τα δίκτυα[1].

Στην παρακάτω ενότητα παρατείθενται μερικές από τις πιο βασικές εφαρμογές των πολύπλοκων δικτύων σε πραγματικά δίκτυα όπως επισης και μαθηματικά αποτελέσματα που αποδεικνύουν το παραπάνω.

1.4 Εφαρμογές πολύπλοκων δικτύων σε πραγματικά δίκτυα.

1.4.1 Internet

Η τοπολογία του Internet μελετάται σε δύο διαφορετικά επίπεδα[8].Στο επίπεδο δρομολογητή που κόμβοι είναι οι δρομολογητές και ακμές είναι οι φυσικές συνδέσεις μεταξύ τους και στο interdomain επίπεδο ή αυτόνομου συστήματος-AS όπου κάθε domain,που αποτελείται από εκατοντάδες δρομολογητές και υπολογιστές παριστάνεται με έναν κόμβο και μία ακμή ενώνει δύο πεδία αν υπάρχει τουλάχιστον μία διαδρομή που τα συνδέει.Το 1999 [9] το Internet μελετήθηκε και στις δύο περιπτώσεις και βρέθηκε ότι η κατανομή βαθμού του δικτύου ακολουθεί power-law.Το 2000 οι Govindan και Tangmunarunkit εξέτασαν τη συνδετικότητα περίπου 200000 δρομολογητών και κατέληξαν στην τιμή 2.3 για τον εκθέτη power-law.

Το Internet σαν δίκτυο παρουσιάζει και ομαδοποίηση και μικρό μήκος μονοπατιού.Το 2001 οι Yook et al. και Pastor-Satorras et al. Μελέτησαν το Internet σε επίπεδο πεδίου και βρήκαν ότι ο συντελεστής ομαδοποίησης κυμαινόταν από 0.18 έως 0.3 σε σύγκριση με αυτόν ενός τυχαίου δικτύου με τις ίδιες παραμέτρους που ήταν 0.001.Βάσει της ίδιας μελέτης,το μέσο μήκος μονοπατιού του Internet σε επίπεδο πεδίου κυμαινόταν από 3.7 έως 3.77 και σε επίπεδο δρομολογητή ήταν περίπου 9,γεγονός που αποδεικνύει τον small-world χαρακτήρα του δικτύου.

Λόγω της ραγδαίας εξάπλωσης των Internet συνδέσεων ανάμεσα στους χρήστες,έκαναν την εμφάνιση τους και άλλα τεχνολογικά δίκτυα που κατά κύριο λόγο έχουν "χτιστεί" πάνω στο ήδη υπάρχων Internet.Τέτοια περίπτωση είναι τα peer-to-peer (p2p) δίκτυα,όπως το Gnutella,που γίνονται όλο και πιο διάσημα,αφού αποτελούν έναν τρόπο διαμοιρασμού αρχείων μεταξύ των χρηστών[7].

1.4.2 World Wide Web(WWW):

Το WWW αποτελεί το μεγαλύτερο δίκτυο του οποίου οι πληροφορίες για την τοπολογία του είναι διαθέσιμες αυτή τη στιγμή [8] .Οι κόμβοι του WWW δικτύου είναι τα έγγραφα (ιστοσελίδες) και οι ακμές είναι οι υπερσύνδεσμοι (URLs) που παραπέμπουν από το ένα έγγραφο στο άλλο.Η κατανομή βαθμού των ιστοσελίδων ακολουθεί power-law.Λόγω του γεγονότος ότι οι ακμές του WWW είναι

κατευθυνόμενες(directed)το δίκτυο χαρακτηρίζεται από δύο κατανομές βαθμού.Την κατανομή των εξερχόμενων ακμων $P_{out}(k)$ που είναι η πιθανότητα ένα έγγραφο να έχει k εξερχόμενες ακμές και την κατανομή εισερχόμενων ακμων $P_{in}(k)$ που είναι η πιθανότητα k υπερσύνδεσμοι να δείχνουν σε ένα συγκεκριμένο έγγραφο.

Πολλές μελέτες έδειξαν ότι και οι δύο κατανομές ακολουθούν power-law δηλαδή ισχύει $P_{out}(k)\sim k^{-\gamma out}$ και $P_{in}(k)\sim k^{-\gamma in}$. Οι Albert,Jeong και Barabasi (1999) μελέτησαν ένα υποσύνολο του δικτύου WWW που περιείχε περίπου 325729 κόμβους και βρήκαν ότι $\gamma_{out}=2.45$ και $\gamma_{in}=2.1$.Οι Kumar et al.(1999) σε μελέτη 40 εκατομμυρίων εγγράφων βρήκαν $\gamma_{out}=2.38$ και $\gamma_{in}=2.1$.Τέλος μία μελέτη σε 200 εκατομμύρια έγγραφα από τους Broder et al.(2000) έδειξε $\gamma_{out}=2.72$ και $\gamma_{in}=2.1$.

Παρατηρώντας τις παραπάνω μελέτες της τοπολογίας του WWW βλέπουμε ότι το γ_{in} παραμένει σταθερό,παρ'όλο τον χρόνο που μεσολάβησε από την πρώτη μέχρι την τελευταία μελέτη,κατά τον οποίο το WWW είχε γίνει τουλάχιστον πενταπλάσιο σε μέγεθος.Αντιθέτως το γ_{out} έχει την τάση να αυξάνεται με την αύξηση του δείγματος ή με το πέρας του χρόνου.

Παρόλο τον μεγάλο αριθμό κόμβων,το WWW παρουσιάζει την ιδιότητα small-world.Το γεγονός αυτό αναφέρθηκε πρώτη φορά από τους Albert,Jeong και Barabasi (1999) που βρήκαν ότι το μέσο μήκος μονοπατιού σε δείγμα 325.729 κόμβων ήταν 11.2 και προέβλεψαν ότι για WWW μεγέθους 800 εκατομμυρίων κόμβων,το μέσο μήκος μονοπατιού θα ήταν περίπου 19.

Η κατευθυνόμενη φύση του WWW δεν επιτρέπει τον υπολογισμό του συντελεστή ομαδοποίησης χρησιμοποιώντας μεθόδους που χρησιμοποιήθηκαν στην περίπτωση του Internet.Ένας τρόπος να αντιμετωπιστεί η δυσκολία αυτή είναι να μετατραπεί το δίκτυο σε μη κατευθυνόμενο.Αυτή τη μέθοδο ακολούθησε η Adamic(1999) σε δείγμα 153.127 ιστοσελίδων και κατέληξε ότι C=0.1078 που είναι πολύ μεγαλύτερο από το $C_{rand}=0.00023$ που αντιστοιχούσε σε τυχαίο γράφο ίδιου μεγέθους και ίδιου βαθμού [7].

1.4.3 Κοινωνικά δίκτυα(social networks):

Ένα κοινωνικο δίκτυο είναι ένα σύνολο ανθρώπων ή κοινωνικών ομάδων που συνδέονται μεταξύ τους με σχέσεις αλληλοεξάρτησης διαφόρων ειδών

(φιλία,ιδέες,συγγένεια,σεξουαλικές επαφές κ.α.).Ο σχηματισμός ενός τέτοιου δικτύου είναι μία πολύπλοκη διαδικασία στην οποία πολλές διαφορετικές οντότητες προσπαθούν ταυτόχρονα να ικανοποιήσουν τους σκοπούς τους.Για παράδειγμα,οι οντότητες αρκετές φορές αλληλεπιδρούν με άλλες,με τις οποίες έχουν κοινά χαρακτηριστικά και προσπαθούν να αποφύγουν σχέσεις με αυτές που έχουν αντικρουώμενα χαρακτηριστικά.

Κατά τη διάρκεια της τελευταίας δεκαετίας αρκετές ομάδες επιστημόνων μελέτησαν τη δομή των κοινωνικών δικτύων μέσα από emails,κινητά τηλέφωνα και εργασιακές σχέσεις.Ουσιαστικά χρησιμοποιήθηκαν δεδομένα από λογαριαμούς email,από αρχεία επικοινωνιών κινητής τηλεφωνίας και από συγκεκριμένες βάσεις δεδομένων με σκοπό την μελέτη και ανάλυση των σχέσεων αλληλεπίδρασης μίας οντότητας.

Τα αποτελέσματα των παραπάνω ερευνών έδειξαν ότι σε όλες τις περιπτώσεις ήταν εμφανές το φαινόμενο small-world όπως επίσης και η συμπεριφορά scale-free[8].Ενδεικτικά αναφέρουμε τα παρακάτω αποτελέσματα.

Σε ένα δίκτυο συνεργασίας ηθοποιών σε ταινίες όπου χρησιμοποιήθηκε η IMDB (Internet Movie Database),το μέσο μήκος μονοπατιού σε δείγμα 225.226 κόμβων (Watts και Strogatz 1998) ήταν 3.65 σε σύγκριση με αυτό ενός τυχαίου γράφου με ίδιες παραμέτρους που ήταν 2.9.Όμως ο συντελεστής ομαδοποίησης ήταν 100 φορές μεγαλύτερος από αυτόν του τυχαίου γράφου.Όσο για την κατανομή βαθμού,αποδείχθηκε ότι ακολουθεί power-law με εκθέτη 2.3+-0.1 (Albert,Barabasi 2000).

Σε ένα δίκτυο email (Ebel,Mielsch,Bornholdt 2002)που χρησιμοποιήθηκε δείγμα 56.969 κόμβων,η κατανομή του βαθμού ακολουθούσε power-law με εκθέτη 1.32+-0.18 και το μέσο μήκος μονοπατιού βρέθηκε 4.95+-0.03 που ήταν πολύ μικρότερο από το αντίστοιχο ενός τυχαίου γράφου με τις ίδιες παραμέτρους L_{rand}=10.10.Τέλος,σχετικά με τον συντελεστή ομαδοποίησης,βρέθηκε ότι $C=3.44*10^{-2}$ που ήταν πολύ μεγαλύτερο από αυτό του τυχαίου γράφου $C_{rand}=4.82*10^{-5}$.[7]

2 Κακόβουλο λογισμικό

Στις μέρες μας,συγκριτικά με προηγούμενα χρόνια,το μεγαλύτερο ποσοστό των υπολογιστών είναι μέλη τουλάχιστον ενός δικτύου.Σε αυτό,συνέβαλε σε μεγάλο βαθμό η εξάπλωση του Internet,αλλά και η τεχνογνωσία σε θέματα δικτύωσης.

Το Internet έχει εισχωρήσει για τα καλά στην ζωή των ανθρώπων και δραστηριότητες που παλιότερα απαιτούσαν την φυσική παρουσία του ατόμου πλέον γίνονται με το πάτημα ενός πλήκτρου.Η ευκολία πρόσβασης σε μεγάλο όγκο πληροφοριών καθώς και η επικοινωνία δίχως χιλιομετρικούς φραγμούς είναι μερικά μόνο από τα πολλά πλεονεκτήματα της χρήσης του Internet.

Παρ'όλα αυτά μαζί με την εξάπλωση της χρήσης του Internet,και την συμμετοχή όλο και περισσότερων υπολογιστών σε δίκτυα,έχουν αυξηθεί και οι απειλές προς αυτούς.Οι απειλές αυτές δεν είναι τίποτα άλλο από το επονομαζόμενο κακόβουλο λογισμικό ή αλλιώς κακόβουλος κώδικας.

Πιο συγκεκριμένα,κακόβουλο λογισμικό είναι το σύνολο εντολών που έχει σχεδιαστεί ώστε να μεταφέρεται από υπολογιστή σε υπολογιστή ή από δίκτυο σε δίκτυο με σκοπό την τροποποίηση των υπολογιστικών ή δικτυακών συστημάτων χωρίς την εξουσιοδότηση του χρήστη ή του ιδιοκτήτη[11].

Όταν ένα κακόβουλο λογισμικό ξεκινάει τη διαδικασία μόλυνσης μεγάλου αριθμού υπολογιστών,λέμε ότι βρίσκεται “in the wild” ή διαφορετικά σε φάση πλήρους ανάπτυξης.Υπάρχουν αυτή τη στιγμή δεκάδες χιλιάδες γνωστά κακόβουλα λογισμικά,που διαχωρίζονται με βάση τον τρόπο επίθεσης και εξάπλωσης τους,από τα οποία ελάχιστα προκαλούν κάποια ανησυχία.Η λίστα wild ή λίστα απειλών είναι ένας

κατάλογος που περιλαμβάνει όλο το κακόβουλο λογισμικό που κυκλοφορεί στο Internet αυτή τη στιγμή και μπορεί κανείς να τον ανατρέξει στη διεύθυνση www.wildlist.org.

Σε αυτό το μέρος του κεφαλαίου θα αναφερθούμε,εν συντομία,στα διάφορα είδη κακόβουλου λογισμικού και στο επόμενο μέρος θα γίνει εκτενής αναφορά στα worms(σκουλήκια) ώστε να είναι εφικτή η μελέτη των μοντέλων εξάπλωσής τους,στο τρίτο και τέταρτο κεφάλαιο.

2.1 Είδη κακόβουλου λογισμικού:

2.1.1 Ιοί (viruses):

Ιός είναι ένα πρόγραμμα το οποίο ξεκινάει τη δράση του σε έναν υπολογιστή χωρίς τη συγκατάθεση του χρήστη αλλά συνήθως απαιτεί κάποια ενέργεια από μέρους του ώστε να εξαπλωθεί.Γενικά οι ιοί αυτο-αντιγράφονται και με αυτόν τον τρόπο εξαπλώνονται από σύστημα σε σύστημα.Πολλοί από αυτούς έχουν το επονομαζόμενο κακόβουλο φορτίο(malicious payload),το οποίο είναι κώδικας που μπορεί να εκτελέσει εντολές στους υπολογιστές-θύματα όπως διαγραφή αρχείων ή απενεργοποίηση του λογισμικού ασφάλειας του υπολογιστή.Επιπλέον πολλοί ιοί έχουν την ικανότητα να προσκολλώνται σε μη κακόβουλα κομμάτια κώδικα ώστε να πραγματοποιούν την εξάπλωση τους.

Οι ιοί γενικά αποτελούνται από τα παρακάτω μέρη:

α) έναν μηχανισμό αντιγραφής που επιτρέπει την αντιγραφή και κατ'επέκταση την εξάπλωση του ιού.

β) έναν μηχανισμό ενεργοποίησης(trigger) που είναι σχεδιασμένος ώστε να ενεργοποιεί τον μηχανισμό αντιγραφής ή να εκτελεί τις διεργασίες του ιού.

γ) ένα καθήκον ή ένα σύνολο καθηκόντων που εκτελούνται σε έναν υπολογιστή με σκοπό την καταστροφή ή αλλαγή αρχείων,την αλλαγή των ρυθμίσεων του υπολογιστή ή γενικότερα την παρεμπόδιση σωστής λειτουργίας του υπολογιστή ή της δικτυωμένης συσκευής.

Οι διαφορετικές μορφές και συμπεριφορές των ιών οφείλονται στα παραπάνω μέρη[12].

Μερικές γνωστές κατηγορίες ιών είναι οι παρακάτω:

- Ο ιός τομέα εκκίνησης(boot sector virus)μολύνει τον πρώτο τομέα μίας δισκέττας ή ενός σκληρού δίσκου.Ο πρώτος τομέας περιέχει την βασική εγγραφή εκκίνησης (master boot record-MBR) όπου βρίσκονται οι οδηγίες και οι πληροφορίες για την εκκίνηση του δίσκου και την εκτέλεση του λειτουργικού συστήματος.Συνεπώς όταν ο υπολογιστής ενεργοποιείται,ο ιός εκτελείται πριν την έναρξη του λειτουργικού συστήματος ή των αντιικών προγραμμάτων πράγμα που καθιστά την ανίχνευση του εξαιρετικά δύσκολη.Άξιο αναφοράς, είναι το γεγονός ότι ο ιός τομέα εκκίνησης δεν εξαπλώνεται μέσω δικτύου σε άλλους υπολογιστές.
- Ο ιός διαγραφής αρχείων(file-deleting virus)έχει για καθήκοντά του την διαγραφή αρχείων με συγκεκριμένη ονομασία και κυρίως αυτών που εκτελούν βασικές λειτουργίες του υπολογιστή ή αυτών που είναι υπεύθυνα για την εκτέλεση εφαρμογών.Μία άλλη κατηγορία τέτοιων ιών είναι αυτοί που στοχεύουν σε τύπους αρχείων όπως εγγραφα κειμένου,λογιστικά φύλλα ή αρχεία γραφικών.
- Ο ιός μόλυνσης αρχείων(file-infecting virus)επισυνάπτει τον εαυτό του σε εκτελέσιμα αρχεία με καταλήξεις .com .exe ή .dll και η διαδικασία εξάπλωσης του ξεκινάει μόλις εκτελεστεί το μολυσμένο αρχείο.Αυτοί οι ιοί είναι παρόμοιοι με τους appender ιούς που επισυνάπτουν αντίγραφο του κώδικά τους στο τέλος ενός αρχείου.Επιπλέον στην ίδια κατηγορία ανήκουν και οι content-embedded ιοί που επισυνάπτονται σε αρχεία γραφικών,html σελίδες,αρχεία βίντεο ή αρχεία ήχου.
- Ο Macro-ιός μπορεί να εξαπλωθεί μέσω macro-εντολών που χρησιμοποιούνται σε εφαρμογές γραφείου όπως το Word και το Excel.Αυτές οι macro-εντολές συνήθως αποθηκεύονται σαν μέρος ενός εγγράφου ή ενός λογιστικού φύλλου και μπορούν να μεταβούν σε άλλα συστήματα όταν τα αρχεία αυτά επισυναφθούν σε ένα μήνυμα email,τοποθετηθούν σε ένα αφαιρούμενο μέσο αποθήκευσης ή αντιγραφούν σε έναν εξυπηρετητή αρχείων ώστε άλλοι χρήστες να έχουν πρόσβαση σε αυτά.
- Ο πολυμορφικός ιός έχει την ικανότητα να παράγει μία μεγάλη ποικιλία από

αντίγραφα του εαυτού του,τα οποία είναι άκρως λειτουργικά.Το γεγονός αυτό καθιστά πολύ δύσκολη την ανίχνευση του από τα αντιικά προγράμματα.

- Ο αόρατος ιός(stealth virus)μπορεί να κρυφτεί από το λειτουργικό σύστημα ή το εκάστοτε αντιικό πρόγραμμα.Αυτό γίνεται εφικτό με το να παρεμποδίζει τα αιτήματα του αντιικού προγράμματος προς στο λειτουργικό σύστημα.Οι αόρατοι ιοί είναι παρόμοιοι με τους αντιευριστικούς ιούς,που οι δημιουργοί κακόβουλου κώδικα σχεδίασαν, για να αποφεύγουν την ευριστική ανίχνευση των αντιικών προγραμμάτων.Ευριστική ανίχνευση είναι η ανίχνευση κατά την οποία αναζητώνται αντιφατικότητες μεταξύ τού κώδικα που ελέγχεται και του κώδικα που πρέπει να έχει κανονικά ένα πρόγραμμα.

2.1.2 Σκουλήκια(worms):

Τα σκουλήκια είναι ένα είδος κακόβουλου λογισμικού το οποίο εγκαθίσταται σε έναν υπολογιστή και στη συνέχεια εξαπλώνεται αυτόματα και χωρίς καμία παρέμβαση από τον χρήστη σε άλλους υπολογιστές μέσω του εκάστοτε δικτύου.

Στο επόμενο μέρος του κεφαλαίου θα ασχοληθούμε πιο διεξοδικά με το συγκεκριμένο είδος κακόβουλου λογισμικού καθώς επίσης θα παραθέσουμε και τις διαφορές του με τους ιούς που συχνά μπερδεύεται.

2.1.3 Πίσω Πόρτες(Backdoors):

Πίσω πόρτα ονομάζουμε ένα είδος κακόβουλου λογισμικού,το οποίο δίνει τη δυνατότητα στον επιτιθέμενο να παρακάμψει τα μέτρα ασφάλειας ενός συστήματος και να αποκτήσει μη εξουσιοδοτημένη πρόσβαση σε αυτό.

Οι πίσω πόρτες μπορεί να είναι είτε αυτόνομα προγράμματα είτε μέρος μη κακόβουλου κώδικα .Παρ'όλα αυτά,δεν έχουν τη δυνατότητα εξάπλωσης σε άλλα συστήματα και άρα η απειλή τους περιορίζεται στο κατά περίπτωση επιτιθέμενο σύστημα.

Η εγκατάσταση ενός τέτοιου κακόβουλου προγράμματος είναι πιθανό να γίνει είτε από τον ίδιο τον κακόβουλο χρήστη,ο οποίος απέκτησε πρόσβαση στο σύστημα

εκμεταλλευόμενος κάποια τρωτά σημεία στην ασφάλεια του,είτε από ιούς ή σκουλήκια τα οποία έχουν ήδη μολύνει το σύστημα και έχουν σαν καθήκον την εγκατάσταση πίσω πόρτας.Σε σπάνιες περιπτώσεις,η εγκατάσταση ενός τέτοιου προγράμματος μπορεί να γίνει ακόμη και από τον ίδιο το χρήστη του συστήματος,εν αγνοία του[12].

2.1.4 Δούρειοι Ίπποι (Trojan Horses):

Ο δούρειος ίππος είναι ένα είδος κακόβουλου λογισμικού παρόμοιο με τις πίσω πόρτες. Οι δούρειοι ίπποι εκ πρώτης όψεως δείχνουν προγράμματα αβλαβή και πολλές φορές χρήσιμα για τον χρήστη του συστήματος.Στην πραγματικότητα όμως πρόκειται για κακόβουλα προγράμματα που ενώ δηλώνουν ότι κάνουν μία νόμιμη διεργασία,αντιθέτως δίνουν τη δυνατότητα στον κακόβουλο χρήστη να έχει πρόσβαση στο σύστημα και εν συνεχεία να υποκλέψει κωδικούς,να κλέψει,να καταστρέψει ή να διαγράψει αρχεία ακόμα και να κάνει προβολή των δραστηριοτήτων του νόμιμου χρήστη σε άλλους υπολογιστές ή δίκτυα.

Οι δούρειοι ίπποι μπορούν να κατηγοριοποιηθούν με βάση τις συνέπειες τους στο μολυσμένο σύστημα σε:

- Απομακρυσμένης πρόσβασης
- Αποστολής δεδομένων
- Καταστροφικούς
- Άρνησης εξυπηρέτησης
- Εξυπηρετητές μεσολάβησης(Proxy Trojan)
- Μεταφοράς,προσθήκης ή διαγραφής αρχείων(FTP Trojan)
- Απενεργοποίησης προγραμμάτων ασφάλειας

Τέλος,άξιο αναφοράς είναι ότι οι δούρειοι ίπποι όπως κι οι πίσω πόρτες,δεν έχουν τη δυνατότητα αυτοματοποιημένης εξάπλωσης σε άλλα συστήματα[12].

2.1.5 Συνδυασμένη απειλή (Blended threat):

Συνδυασμένη απειλή είναι το είδος κακόβουλου λογισμικού που μπορεί να

αντιγράψει τον εαυτό του με περισσότερους από έναν τρόπους,μπορεί να έχει περισσότερους από έναν μηχανισμούς ενεργοποίησης και μπορεί να έχει την ικανότητα διεκπεραίωσης πολλαπλών καθηκόντων.Ένα τέτοιο είδος κακόβουλου λογισμικού είναι συχνά ικανό να εξαπλώνεται στο δίκτυο ακριβώς όπως ένα σκουλήκι.Επιπλεόν,κατά την επίθεσή του μπορεί να τοποθετήσει έναν δούρειο ίππο στον υπολογιστή.Ουσιαστικά το συγκεκριμένο είδος κακόβουλου λογισμικού,όπως προδίδει και το όνομα του,είναι ένας συνδυασμός χαρακτηριστικών και ικανοτήτων πολλών διαφορετικών απειλών[12].

2.1.6 Ωρολογιακές Βόμβες (Time Bombs):

Η ωρολογιακή βόμβα(ή αλλιώς λογική βόμβα-logic bomb)αποτελεί μία από τις πρώτες μορφές κακόβουλου λογισμικού.Όταν εγκατασταθεί μία ωρολογιακή βόμβα,παραμένει αδρανής μέχρι να ενεργοποιηθεί από κάποιο ερέθισμα που έχει ορίσει ο δημιουργός της.Το συγκεκριμένο ερέθισμα μπορεί να είναι είτε κάποια συγκεκριμένη ημερομηνία,είτε συγκεκριμένη ώρα,είτε συγκεκριμένος αριθμός επανεκκινήσεων του συστήματος.

Μία ωρολογιακή βόμβα δεν έχει τη δυνατότητα αυτοματοποιημένης εξάπλωσης σε άλλα συστήματα[12].

2.1.7 Λογισμικό υποκλοπής (Spyware):

Τον όρο spyware τον χρησιμοποιούμε για να περιγράψουμε οποιοδήποτε κακόβουλο λογισμικό συλλέγει πληροφορίες για τον χρήστη του συστήματος χωρίς αυτός να το γνωρίζει και χωρίς να έχει δώσει τη συγκατάθεσή του.Η εγκατάσταση ενός spyware προγράμματος μπορεί να είναι είτε επακόλουθο μόλυνσης του συστήματος από ιό,είτε επακόλουθο εγκατάστασης ενός καινούριου προγράμματος. Οι πληροφορίες που συνήθως συλλέγονται από προγράμματα spyware είναι λίστες από ιστοσελίδες που επισκεπτεται συχνότερα κάποιος χρήστης,πληροφορίες για τις εφαρμογές ή το λειτουργικό σύστημα που είναι εγκατεστημένα,αλλά τις περισσότερες φορές οι πληροφορίες που συλλέγονται είναι πιο ευαίσθητες και αφορούν κωδικούς πρόσβασης,ονόματα χρήστη,ακόμα και αριθμούς πιστωτικών καρτών.Οι

δραστηριότητες του χρήστη-θύματος καταγράφονται,οι πληροφορίες συλλέγονται και στη συνέχεια αποστέλλονται στο κακόβουλο χρήστη[12].

2.1.8 Adware:

Το adware είναι το λογισμικό που ευθύνεται για την εμφάνιση διαφημίσεων στον υπολογιστή ενός χρήστη.Η λειτουργία του είναι παρόμοια με αυτή του spyware.Πληροφορίες σχετικές με τις προτιμήσεις και τις συνήθειες ενός Internet χρήστη,συλλέγονται,τοποθετούνται σε βάσεις δεδομένων και τελικά επιλέγονται οι διαφημίσεις που θα προβληθούν στον υπολογιστή του,ώστε να συμβαδίζουν με τις προτιμήσεις του.Οι ιστοσελίδες που χρησιμοποιούν κώδικα adware ισχυρίζονται ότι με αυτόν τον τρόπο βελτιώνεται η εξυπηρέτηση των πελατών και κατ'επέκταση αποφέρει χρήματα στην ιστοσελίδα.Παρ'ότι η συγκεκριμένη διαδικασία φαίνεται αβλαβής,το επικίνδυνο είναι ο τρόπος διαχείρισης των πληροφοριών που συλλέχθησαν καθώς και τα άτομα που έχουν πρόσβαση στις συγκεκριμένες πληροφορίες[12].

2.2 Σκουλήκια (Worms):

Όπως αναφέραμε εν συντομία στο πρώτος μέρος του κεφαλαίου,τα σκουλήκια είναι ένα είδος κακόβουλου λογισμικού το οποίο εγκαθίσταται σε ένα σύστημα και στη συνέχεια εξαπλώνεται αυτόματα και χωρίς καμία παρέμβαση από τον χρήστη σε άλλα συστήματα χρησιμοποιώντας το δίκτυο.

Σε πολλές περιπτώσεις τα σκουλήκια μπερδεύονται με τους ιούς λόγω αρκετών κοινών τους χαρακτηριστικών.Η παραπάνω σύγχυση οδηγεί τις περισσότερες φορές στην ανεπαρκή προστασία ενός συστήματος,καθώς δεν λαμβάνονται υπόψιν σημαντικά χαρακτηριστικά των σκουληκιών που δεν υπάρχουν στους ιούς.

Μερικές βασικές διαφορές που παρατηρούνται ανάμεσα στα σκουλήκια και τους ιούς ακολουθούν παρακάτω:

- Τόσο τα σκουλήκια όσο και οι ιοί εξαπλώνονται σε άλλα συστήματα.Παρ'όλα

αυτά,οι ιοί συνήθως εξαπλώνονται προσκολλώντας τον εαυτό τους σε άλλα αρχεία.Συνεπώς βασική προυπόθεση για την εξάπλωση ενός ιού,είναι η μετάδοση του μολυσμένου αρχείου.Αντιθέτως τα σκουλήκια,έχουν τη δυνατότητα να εξαπλώνονται αυτόνομα και αυτόματα από σύστημα σε σύστημα μέσω του δικτύου.

- Ένα σκουλήκι έχει την ικανότητα να ελέγξει και να χρησιμοποιήσει το δίκτυο για να ανιχνεύσει και στη συνέχεια να μολύνει το σύστημα-στόχο.Αντιθέτως ένας ιός δεν έχει τις προαναφερθείσες ικανότητες.

Από το 1998 με το σκουλήκι Morris μέχρι και σήμερα,τα Internet σκουλήκια έχουν προκαλέσει τις μεγαλύτερες σε μέγεθος ζημιές από όλα τα είδη κακόβουλου λογισμικού για υπολογιστές.Αναφέρουμε για μία ακόμη φορά ότι το Internet σκουλήκι είναι ένα κομμάτι κακόβουλου κώδικα που αντιγράφεται και εξαπλώνεται μέσω συνδέσεων δικτύου,χωρίς να απαιτείται κάποια παρέμβαση από τον χρήστη-θύμα.

Ο κύκλος ζωής ενός σκουληκιού,μετά την απελευθέρωση από τον δημιουργό του,αποτελείται από τις εξής φάσεις:εύρεση στόχου,μετάδοση σκουληκιού, ενεργοποίηση σκουληκιού και τελικά μόλυνση του στόχου.

Κατά τη διάρκεια εύρεσης στόχου και μετάδοσης,το σκουλήκι είναι ενεργό στο δίκτυο γεγονός που καθιστά την ανίχνευσή του δυνατή,από τα συστήματα ανίχνευσης εισβολών(NIDS-Network-based intrusion detection systems).

Οι δραστηριότητες των δύο τελευταίων φάσεων περιορίζονται σε τοπικά πλαίσια και συνεπώς είναι δυσκολότερη η ανιχνευση από τα NIDSs.

Παρακάτω θα κατηγοριοποιήσουμε τα σκουλήκια με βάση τους μηχανισμούς εύρεσης στόχου,διάδοσης,μετάδοσης και μορφής φορτίου.

2.2.1 Μηχανισμοί εύρεσης στόχου:

Το πρώτο στάδιο στη ζωή ενός σκουληκιού είναι η εύρεση στόχων,δηλαδή η σάρωση του δικτύου για ευάλωτα συστημάτα που μπορούν να μολυνθούν.Υπάρχουν

πολλοί διαφορετικοί μηχανισμοί που μπορούν να χρησιμοποιηθούν από ένα σκουλήκι για να ανιχνεύσει το επόμενο σύστημα-θύμα οι οποίοι παρατίθενται παρακάτω[13].

2.2.1.1 *Τυχαία σάρωση (Random scan):*

Τυχαία σάρωση ονομάζουμε τον μηχανισμό σάρωσης που χρησιμοποιεί ένα σκουλήκι κατά την οποία σαρώνει ολόκληρο το διάστημα διευθύνσεων IP,όπου περιέχονται 2^{32} διαφορετικές IP διευθύνσεις,στην περίπτωση του IPv4.

2.2.1.2 *Επιλεκτική τυχαία σάρωση (Selective Random scan):*

Σε αντίθεση με την τυχαία σάρωση,ένα σκουλήκι μπορεί να σαρώσει ένα μέρος ,του διαστήματος διευθύνσεων IPv4, που είναι πιο πιθανό να χρησιμοποιείται στο Internet.Το γεγονός αυτό θα βοηθήσει το σκουλήκι να εξαπλωθεί γρηγορότερα αφού δεν θα σπαταληθεί χρόνος για τη σάρωση IP διευθύνσεων που δεν έχουν ανατεθεί.Για την επιλογή του μέρους του διαστήματος IP διευθύνσεων μπορεί να χρησιμοποιηθεί ο χάρτης ανάθεσης IPv4 διευθύνσεων του οργανισμού IANA.Παρ'όλο που αρκετά σκουλήκια (π.χ. Slapper) έχουν χρησιμοποιήσει τη συγκεκριμένη μέθοδο σάρωσης για να αυξήσον τη ταχύτητα εξάπλωσης τους,πρέπει να σημειωθεί ότι το μέγεθος του κώδικα ενός τέτοιου σκουληκιού αυξάνεται δραματικά.Ο λόγος που συμβαίνει αυτό είναι επειδή το σκουλήκι πρέπει να κουβαλάει τις πληροφορίες που σχετίζονται με τις επιλεγμένες IP διευθύνσεις που θα στοχεύσει η σάρωση.Συνεπώς όσο περισσότερες πληροφορίες κουβαλάει ένα σκουλήκι,τόσο αυξάνεται το μέγεθος του κώδικά του και επιβραδύνεται η εξάπλωσή του.

2.2.1.3 *Σειριακή σάρωση με τοπική προτίμηση(Sequential scan with local preference):*

Κατά τη διαδικασία σειριακής σάρωσης,το σκουλήκι αρχίζει να σαρώνει IP διευθύνσεις.Το σκουλήκι θα επιλέξει με μεγαλύτερη πιθανότητα μία διεύθυνση κοντά στη δική του για να ξεκινήσει τη σάρωση,παρά κάποια μακρινή.Όταν κάποιο μολυσμένο τερματικό εντοπίσει και μολύνει ένα ευάλωτο τερματικό που έχει διεύθυνση έστω x,το αρχικά μολυσμένο τερματικό θα συνεχίσει με τη σάρωση των IP διευθύνσεων x+1,x+2,... .Εν συνεχεία,το ευάλωτο τερματικό που μολύνθηκε,θα

διαλέξει με μεγάλη πιθανότητα την αρχική διεύθυνση σάρωσης του αρχικά μολυσμένου τερματικού και κατ'επέκταση την ίδια διαδρομή σάρωσης.Είναι προφανές,ότι η πιθανότητα να σαρώνονται συνεχώς τα ίδια τερματικά είναι μεγάλη,κάτι που επιβραδύνει δραματικά την εξάπλωση του σκουληκιού.

2.2.1.4 *Routable scan*

Παρ'όλο που τα περισσότερα σκουλήκια σαρώνουν στα τυφλά ολόκληρο το διάστημα διευθύνσεων IPv4 ένα σκουλήκι που χρησιμοποιεί το μηχανισμό routable σάρωσης (routing σκουλήκι) εστιάζει σε ένα μικρότερο διάστημα σάρωσης.Το συγκεκριμένο είδος σκουληκιού χρησιμοποιεί τις πληροφορίες που παρέχονται από το πρωτόκολλο εξωτερικής δρομολόγησης (Border gateway protocol-BGP) ώστε να περιορίσει το φάσμα σάρωσης και να στοχεύσει σε συγκεκριμένα συστήματα εντός μιας γεωγραφικής περιοχής,έναν πάροχο υπηρεσιών Internet ή ένα αυτόνομο σύστημα.Ένα τέτοιο είδος σκουληκιού έχει την ικανότητα να εξαπλώνεται ακόμα και τρεις φορές γρηγορότερα από ένα σκουλήκι που χρησιμοποιεί τυχαία σάρωση.

2.2.1.5 *Hit-List scan*

Σε αντίθεση με τις προηγούμενες μεθόδους που θεωρούνται “τυφλές” ,γιατί το σκουλήκι δεν έχει προηγούμενη γνώση για τους ευάλωτους στόχους,η μέθοδος σάρωσης με βάση μία λίστα στόχων χρησιμοποιεί έναν ήδη υπάρχοντα κατάλογο ευάλωτων διευθύνσεων.Ο κατάλογος αυτός,προκύπτει από σάρωση του Internet,με σκοπό την εύρεση συστημάτων με τρωτά σημεία απέναντι σε ένα συγκεκριμένο σκουλήκι,τα οποία συστήματα καταγράφονται μαζί με τις διευθύνσεις τους.Με αυτόν τον τρόπο,το σκουλήκι γνωρίζει ακριβώς που βρίσκεται ο στόχος.Η προαναφερθείσα λίστα συνήθως δημιουργείται πριν απελευθερωθεί το σκουλήκι.Επίσης,μπορεί να περιλαμβάνεται εντός του κακόβουλου κώδικα ή να είναι αποθηκευμένη κάπου εξωτερικά,ώστε το οποιοδήποτε σκουλήκι να έχει πρόσβαση.

Είναι προφανές ότι όσο μεγαλύτερη είναι η λίστα στόχων,τόσο πιο δύσκολα μπορεί να κουβαληθεί ή να αποκτηθεί από το σκουλήκι.Παρ'όλα αυτά,με μία αρκετά μεγάλη λίστα,η σάρωση είναι πιο ακριβής και συνεπώς το σκουλήκι μπορεί να προκαλέσει μεγαλύτερη ζημιά.

Παράδειγμα ενός γνωστού σκουληκιού που χρησιμοποίησε τον συγκεκριμένο μηχανισμό εύρεσης στόχων είναι το σκουλήκι Warhol το οποίο ήταν ικανό να μολύνει σχεδόν όλα τα ευάλωτα συστήματα σε λιγότερο από 15 λεπτά.

2.2.1.6 *Παθητική προσέγγιση (Passive approach):*

Βάσει του συγκεκριμένου μηχανισμού ανίχνευσης στόχων,το σκουλήκι δεν ψάχνει για ευάλωτα συστήματα,παρά περιμένει υπομονετικά τα πιθανά θύματα να προσεγγίσουν το ήδη μολυσμένο σύστημα στο οποίο βρίσκεται.Είναι πιθανό επίσης να περιμένει συγκεκριμένες ενέργειες από τον χρήστη ώστε να βρεθεί το επόμενο θύμα.

Ένα παράδειγμα της τελευταίας περίπτωσης είναι το παθητικό σκουλήκι Gnuman που λειτουργει σαν κόμβος Gnutella και περιμένει queries ώστε να αντιγράψει τον εαυτό του.Η συγκεκριμένη μέθοδος είναι αργή,αλλά αντιμετωπίζεται πολύ δύσκολα.

2.2.2 Μηχανισμοι διάδοσης σκουληκιού:

Μετά την διαδικασία εύρεσης του επόμενου θύματος,ένα αντίγραφο του σκουληκιού στέλνεται στον στόχο.Οι τρεις διαφορετικοί μηχανισμοί διάδοσης ενός σκουληκιού είναι:self-carried,second channel και embedded[13].

Στα self-carried σκουλήκια,η διάδοση είναι απευθείας.Δηλαδή,το φορτίο το σκουληκιού μεταφέρεται από μόνο του σε ένα πακέτο.

Άλλα σκουλήκια διαδίδονται μέσω second channel(δεύτερου καναλιού)που σημαίνει ότι αφού βρεθεί ο στόχος,το σκουλήκι πάει αρχικά στο στόχο και στη συνέχεια κατεβάζει το κακόβουλο φορτίο από το Internet ή από ένα ήδη μολυσμένο σύστημα με χρήση μίας πίσω πόρτας,η οποία έχει εγκατασταθεί μέσω απομακρυσμένης κλήσης διαδικασιών(Remote procedure call-RPC) ή άλλων εφαρμογών.

Ο μηχανισμός embedded (ενσωματωμένης) διάδοσης θεωρείται ο "αόρατος" μηχανισμός.Το σκουλήκι που χρησιμοποιεί τον συγκεκριμένο μηχανισμό δύσκολα

ανιχνεύεται από τα anomally-based συστήματα ανίχνευσης εισβολής αφού κατά τη διαδικασία διάδοσης δεν δημιουργείται καμία ανωμαλία στις διεργασιες του συστήματος που μολύνεται.

Επιπλέον στους μηχανισμούς που προαναφέρθηκαν,προστίθενται και τα botnets (δίκτυα ρομπότ) που έχουν χρησιμοποιηθεί για διάδοση σκουληκιών, spyware ,spam και εκτέλεση επίθεσης κατανεμημένης άρνησης εξυπηρέτησης(DDoS).

2.2.3 Μηχανισμοί μετάδοσης σκουληκιών:

Βάσει του τρόπου που μεταδίδονται τα σκουλήκια ,διαχωρίζονται σε TCP και UDP σκουλήκια.Η κύρια διαφορά μεταξύ αυτών των δύο ειδών είναι ότι τα TCP σκουλήκια έιναι latency-limited(περιορισμένα από την καθυστέρηση) ενώ τα UDP σκουλήκια είναι bandwidth-limited(περιορισμένα από το εύρος ζώνης)[13].Όλες οι TCP συνδέσεις απαιτούν χειραψία τριών δρόμων(three-way handshake)για να εδραιώσουν σύνδεση πριν τη μετάδοση.Συνεπώς όταν ένα τερματικό στέλνει ένα TCP SYN πακέτο για την έναρξη της σύνδεσης,πρέπει να περιμένει μέχρι να λάβει το αντίστοιχο SYN/ACK ή timeout πακέτο από το άλλο άκρο,πρωτού κάνει οποιαδήποτε ενέργεια.Συγκριτικά με τα UDP σκουλήκια,τα TCP χρειάζονται ένα επιπλέον RTT(round-trip time)και δύο πακέτα των 40-bytes για την εδραίωση της σύνδεσης.Κατά τη διάρκεια του συγκεκριμένου χρόνου αναμονής η διαδικασία μόλυνσης έχει μπλοκαριστεί και δεν μπορούν να μολυνθούν άλλα τερματικά.

Τα UDP σκουλήκια δεν χρειάζονται την εδραίωση της σύνδεσης για να ξεκινήσει η διαδικασία της μόλυνσης λόγω της φύσης του UDP.Στην περίπτωση αυτή,το σκουλήκι είναι Self-carried και περιέχεται στο πρώτο πακέτο που στέλνεται στον στόχο.Από τη στιγμή που δεν υπάρχει χρόνος αναμονής ,όπως στα TCP σκουλήκια,τα UDP εξαπλώνονται πολύ γρηγορότερα και η ταχύτητα τους περιορίζεται μόνο από το εύρος ζώνης.Για αυτό το λόγο τα UDP σκουλήκια συχνά ανταγωνίζονται το ένα το άλλο για τους πόρους του δικτύου.

2.2.4 Μορφή φορτίου σκουληκιού:

Με τον όρο φορτίο εννοούμε τον κώδικα του σκουληκιού.Είναι συνηθισμένο τα σκουλήκια να στέλνουν το φορτίο τους απευθείας.Όμως τα signature-based συστήματα ανίχνευσης αντιστοιχίζουν το φορτίο του σκουληκιού με τις υπογραφές στη βάση δεδομένων και συνεπώς γίνεται η ανίχνευση του σκουληκιού.Ορισμένα σκουλήκια διαφοροποιούν το μέγεθος του φορτίου τους τοποθετώντας άχρηστα δεδομένα μέσα στον κώδικα.Ακόμα και έτσι όμως,η υπογραφή το σκουληκιού δεν αλλάζει ώστε να αποτραπεί η ανίχνευσή του.Αυτά τα σκουλήκια λέγονται μονομορφικά[13].

Οι δημιουργοί των σκουληκιών μπορούν να κάνουν αλλαγές στο φορτίο,κάνοντας το σκουλήκι να μοιάζει αβλαβές ώστε να αποφύγει τα συστήματα ανίχνευσης.

Ο όρος πολυμορφικό χρησιμοποιείται για να περιγράψει τα σκουλήκια που αλλάζουν το φορτίο τους δυναμικά,έτσι ώστε σε κάθε χρονική στιγμή το σκουλήκι να μοιάζει διαφορετικό αλλά να λειτουργεί ακριβώς με τον ίδιο τρόπο.Η αναγνώριση από ένα παραδοσιακό σύστημα ανίχνευσης είναι πολύ δύσκολη όταν το σκουλήκι αλλάζει συνεχώς εμφάνιση.

Τελευταία κατηγορία σκουληκιών βάσει της μορφής του φορτίου τους είναι τα μεταμορφικά σκουλήκια.Τα συγκεκριμένα σκουλήκια εκτός από την εμφάνιση τους αλλάζουν και τη λειτουργία του κώδικά τους σε τέτοιο βαθμό,ώστε να αποφευχθεί η ανίχνευση τους αλλα να μην χαθεί η δραστικότητά τους.ενώ αν σε όλα αυτά προστεθεί και ένας πολύπλοκος μηχανισμός κρυπτογράφησης ώστε να μην είναι εμφανής ο πραγματικός σκοπός τους,τότε η άμυνα απέναντί τους είναι ακόμα δυσκολότερη.

2.2.5 "Διάσημα" Internet σκουλήκια:

***2.2.5.1** The Morris Worm:*

Το σκουλήκι Morris ήταν ένα από τα πρώτα Internet σκουλήκια που έγινε

ευρέως γνωστό λόγω των καταστρεπτικών συνεπειών του.Δημιουργήθηκε από τον Robert Tappan Morris, φοιτητή του πανεπιστημίου Cornell και απελευθερώθηκε τον Νοέμβριο του 1988.Σύμφωνα με τον δημιουργό του,το σκουλήκι Morris δεν σχεδιάστηκε με κακές προθέσεις,αλλά για να ανακαλύψει τον αριθμό των τερματικών στο Internet.Το σκουλήκι είχε σχεδιαστεί για να τρέχει μία διαδικασία σε κάθε μολυσμένο τερματικό,η οποία έδινε απάντηση στην ερώτηση αν το συγκεκριμένο τερματικό είναι μολυσμένο από το σκουλήκι Morris.Αν η απάντηση ήταν "ναι" το ήδη μολυσμένο τερματικό θα έπρεπε να προσπεραστεί,αν η απάντηση ήταν "όχι"το σκουλήκι αντέγραφε τον εαυτό του στο τερματικό.Στην πραγματικότητα,ένα λάθος στον κώδικα,ανάγκαζε το σκουλήκι να αντιγράφει τον εαυτό του πολλές φορές σε ήδη μολυσμένες συσκευές,τρέχοντας κάθε φορά μία νέα διαδικασία.

Σαν αποτέλεσμα,η διαθέσιμη επεξεργαστική ισχύς των τερματικών που μολύνονταν μειωνόταν συνεχώς,με αποτέλεσμα τα τερματικά να καθιστώνται μη χρηστικά[14].

2.2.5.2 *Code Red I και Code Red II:*

Το Code Red I εμφανίστηκε για πρώτη φορά τον Ιούλιο του 2001 απειλώντας υπολογιστές που έτρεχαν Microsoft Internet Information Server (IIS) Web Service.Τις πρώτες 20-25 μέρες,αφότου μόλυνε έναν υπολογιστή,το Code Red I χρησιμοποιούσε έναν "τυφλό" μηχανισμό σάρωσης που σάρωνε το port 80 σε τυχαίες IP διευθύνσεις,ώστε να βρεί άλλα ευάλωτα τερματικά.Οι μολυσμένες ιστοσελίδες εμφάνιζαν το μήνυμα "HELLO!Welcome to http://www.worm.com!Hacked by Chinese!"

Το Code Red II απελευθερώθηκε ένα μήνα μετά.Πρόκειται για μία παραλλαγή του Code Red I, η οποία δεν εκτελούσε πλέον επίθεση DoS σε προκαθορισμένες IP διευθύνσεις.Αντ'αυτού εγκαθιστούσε μία πίσω πόρτα στα μολυσμένα συστήματα.Όπως και ο προκάτοχος του έτσι και το Code Red II χρησιμοποιεί ¨τυφλή" σάρωση με τη διαφορά ότι εστιάζει κυρίως στο τοπικό υποδίκτυο και στοχεύει σε συστήματα με κινέζικη ρύθμιση γλώσσας.

Το φορτίο του Code Red I είναι μονομορφικό και η υπογραφή του ξεκινά με "GET/default.ida?NNNNNNN.".Το Code Red II έχει παρόμοια υπογραφή με τον

προκάτοχό του μόνο που αντί για N έχει X.

Και οι δύο εκδόσεις του Code Red είναι self-carried και μεταδίδονται μέσω TCP συνδέσεων[13].

2.2.5.3 *Slammer/Sapphire:*

Το σκουλήκι Slammer γνωστό και ως Sapphire είναι ένα από τα μικρότερα σε μέγεθος σκουλήκια που έχουν εμφανιστεί.Βρέθηκε τον Ιανουάριο του 2003 και απειλούσε συστήματα με Microsoft SQL Server 2000 ή MSDE 2000.

Το Slammer χρησιμοποιεί UDP port 1434 για να εκμεταλλευθεί μία υπερχείλιση buffer σε έναν MS SQL Server.Το μέγεθος του κώδικα είναι 376bytes και προσθέτοντας την κεφαλίδα UDP φτανει τα 404 bytes στο σύνολο[15].

Χρησιμοποιεί έναν μηχανισμό "τυφλής" σάρωσης,όπου αριθμοί παράγονται τυχαία και χρησιμοποιούνται σαν IP διευθύνσεις για την διαδικασία ανεύρεσης ευάλωτων τερματικών.Για να ξεκινήσει η γεννήτρια τυχαίων αριθμών,το Slammer χρησιμοποιεί τη συνάρτηση GetTickCount() από το Win32API.Μερικές φορές η γεννήτρια είναι πιθανό να επιστρέψει IP διευθύνσεις της μορφής α.β.γ.255 που είναι διευθύνσεις broadcast .Έτσι το Slammer εξαπλώνεται γρηγορότερα αφού όλα τα τερματικά του συγκεκριμένου δικτύου θα μολυνθούν.

Τέλος,όπως και τα περισσότερα σκουλήκια UDP έτσι και το Slammer είναι self-carried και έχει μονομορφικό φορτίο.

3 Επιδημιολογικά μοντέλα και λοιπά μοντέλα εξάπλωσης σκουληκιών:

Οι ιοί και τα σκουλήκια στους υπολογιστές είναι παρόμοιοι με τους ιούς στη βιολογία,τόσο στις συμπεριφορές που σχετίζονται με την αντιγραφή του εαυτού τους,όσο και στις συμπεριφορές που σχετίζονται με τη εξάπλωση τους.Συνεπώς τα μαθηματικά μοντέλα που αναπτύχθηκαν για τη μελέτη βιολογικών μολυσματικών ασθενειών μπορούν να προσαρμοστούν έτσι ώστε να χρησιμοποιηθούν και στη μελέτη διάδοσης κακόβουλου λογισμικού.Στην παρούσα εργασία θα ασχοληθούμε

αποκλειστικά με τα μοντέλα εξάπλωσης σκουληκιών.

Στην επιδημιολογία γενικά,υπάρχουν δύο κατηγορίες μοντέλων για την περιγραφή εξάπλωσης μολυσματικών ασθενειών ή κακόβουλου λογισμικού (στην περίπτωση μας).Τα στοχαστικά μοντέλα,που είναι κατάλληλα για την μελέτη εξάπλωσης σε μικρής κλίμακας δίκτυα και τα ντετερμινιστικά μοντέλα που είναι κατάλληλα για την μελέτη εξάπλωσης σε μεγάλης κλίμακας δίκτυα.

Τα επιδημιολογικά μοντέλα χαρακτηρίζονται από δύο σημαντικές προτάσεις:

α) Σε δεδομένο χρόνο t κάθε κόμβος του δικτύου μπορεί να βρίσκεται σε μία μόνο κατάσταση(από έναν περιορισμένο αριθμό καταστάσεων) όπως για παράδειγμα ευάλωτος,μολυσματικός,απομακρυσμένος κλπ.Η επιλογή των καταστάσεων που θα περιέχονται στο μοντέλο εξαρτάται από τα χαρακτηριστικά του εκάστοτε σκουληκιού που αναλύεται και από τον σκοπό του μοντέλου.

β) Ο μηχανισμός μετάδοσης ενός σκουληκιού περιγράφεται με την πιθανότητα ένας κόμβος να μολύνει κάποιον άλλο.Κατά τον ίδιο τρόπο,οι μεταβάσεις των κόμβων από κατάσταση σε κατάσταση,περιγράφονται από απλές πιθανότητες.

Στα παρακάτω επιδημιολογικά μοντέλα θεωρούμε ότι η διάδοση λαμβάνει χώρα σε έναν γράφο με N κόμβους και m ακμές.Με S(t) παριστάνουμε τον αριθμό των ευάλωτων κόμβων σε χρόνο t,με I(t) τον αριθμό των μολυσματικών κόμβων και με R(t) τον αριθμό των κόμβων που έχουν απομακρυνθεί.Σε όλες τις περιπτώσεις με β παριστάνεται ο ρυθμός με τον οποίο οι ευάλωτοι κόμβοι μολύνονται.Σε περίπτωση που χρησιμοποιηθούν άλλοι συμβολισμοί,θα γίνεται αναφορά στην εκάστοτε περίπτωση.

Στα περισσότερα μοντέλα εξάπλωσης σκουληκιών,γίνεται η υπόθεση ότι το β είναι σταθερό,εξισταθμίζοντας με αυτόν τον τρόπο τις διαφορές σε επεξεργαστική ισχύ,εύρος ζώνης δικτύου και τοποθεσία του μολυσματικού κόμβου.

3.1 Μοντέλο SI (Susceptible-Infectious model):

Σε αυτή την κατηγορία μοντέλων,από τη στιγμή που ένας ευάλωτος κόμβος γίνει μολυσματικός,η κατάσταση του παραμένει μολυσματική καθόλη τη διάρκεια της εξάπλωσης του σκουληκιού.Αυτά τα μοντέλα μπορούν να χρησιμοποιηθούν για τη μελέτη της χειρότερης περίπτωσης μίας διάδοσης,όταν δεν υπάρχουν

αυτοματοποιημένα και ανθρώπινα αντίμετρα.

Έστω d ο μέσος βαθμός του δικτύου και i(t) η αναλογία μολυσματικών κόμβων στο δίκτυο για χρόνο t.Ο αναμενόμενος αριθμός ευάλωτων γειτόνων που μπορεί να μολυνθεί από έναν συγκεκριμένο μολυσματικό κόμβο είναι d(1-i(t)).Από τη στιγμή που υπάρχουν I(t) μολυσματικοί κόμβοι συνολικά,ο συνολικός ρυθμός με τον οποίο προκύπτουν καινούριοι μολυσμένοι κόμβοι στο δίκτυο είναι βd(1-i(t))i(t).Συνεπώς το γενικό μοντέλο SI περιγράφεται από την παρακάτω διαφορική εξίσωση (1)

$$\frac{di(t)}{dt} = \beta d(1-i(t))i(t)$$

με αρχικές εξισώσεις

$$i(0) = \frac{I(0)}{N} > 0$$

και

$$\forall t \geq 0, i(t) + s(t) = 1$$

Η λύση της (1) που περιγράφει την αναλογία των μολυσματικών κόμβων είναι η λογιστική καμπύλη(η πιο κοινή σιγμοειδής καμπύλη)

$$i(t) = \frac{i(0)e^{\beta' t}}{1 - i(0) + i(0)e^{\beta' t}}$$

όπου β' = βd

Η παραπάνω σιγμοειδής καμπύλη μπορεί να χωριστεί σε τρεις περιοχές[16]:

α) αργή εκκίνηση,όπου ελάχιστοι κόμβοι μολύνονται σε κάθε βήμα χρόνου.

β) εκθετική αύξηση,όπου ο αριθμός των κόμβων που μόλις μολύνθηκαν αυξάνεται εκθετικά.

γ) κατάσταση ισορροπίας,όπου ο αριθμός των μολυσματικών κόμβων παίρνει κάποια μέγιστη τιμή γύρω από την οποία ταλαντώνεται σταθερά.

Άν το σκουλήκι διαδίδεται σε πλήρη γράφο N κόμβων όπου d=N-1,η

διαφορική εξίσωση (1) μπορεί να γραφεί ως εξής(2):

$$\frac{di(t)}{dt} = \beta(1-i(t))I(t)$$

με αρχικές εξισώσεις

$$i(0) = \frac{I(0)}{N} > 0$$

και

$$\forall t \geq 0, i(t) + s(t) = 1$$

άρα

$$i(t) = \frac{i(0)e^{\beta(N-1)t}}{1-i(0)+i(0)e^{\beta(N-1)t}}$$

Οι Staniford,Paxson κ Weaver[17]χρησιμοποίησαν το συγκεκριμένο μοντέλο(2)για την μελέτη της εξάπλωσης του σκουληκιού Code Red I.Προσαρμόζοντας το παραπάνω μοντέλο στα δεδομένα που είχαν συλλέξει οι Chemical Abstract Services εκτίμησαν ότι το γινόμενο β(N-1) ήταν 1.8.Παρόλα αυτά,τα αποτελέσματα της μελέτης τους,θεωρήθηκαν λανθασμένα αφού για την παραπάνω εκτίμηση χρησιμοποίησαν τον αριθμό των κόμβων που είχαν σαρωθεί,ο οποίος ήταν πολύ μεγαλύτερος του αριθμού των μολυσματικών κόμβων.

Ο Weaver[18]και ο Wagner κι η ομάδα του [19] χρησιμοποίησαν το μοντέλο αυτό για να μελετήσουν τέσσερις μηχανισμούς τοπικής διάδοσης: hitlist,τοπολογικό,μεταθετικό και τοπικό υποδίκτυο.Όμως ο πλήρης γράφος,που πήραν σαν τοπολογικό δεδομένο οι παραπάνω,είναι τοπολογικά ακατάλληλος για την μελέτη τέτοιων τοπικών μηχανισμών.

Τέλος αξίζει να σημειωθεί ότι θεωρώντας τοπολογία τυχαίου γράφου ER (Erdos- Renyi)με πυκνότητα ακμών p,ο βαθμός ενός κόμβου είναι p(N-1) και το μοντέλο SI που περιγράφει την διάδοση σε αυτού του είδους τα δίκτυα,περιγράφεται από τη διαφορική εξίσωση

$$\frac{di(t)}{dt} = \beta p(N-1)(1-i(t))i(t)$$

με λύση την

$$i(t) = \frac{i(0)e^{\beta p(N-1)t}}{1-i(0)+i(0)e^{\beta p(N-1)t}}$$

3.2 Μοντέλο SIS (Susceptible-Infectious-Susceptible):

Σε αυτή τη κατηγορία μοντέλων,ένας μολυσματικός κόμβος αναρρώνει και συνεπώς γίνεται ευάλωτος ξανά.Αυτά τα μοντέλα μπορούν να χρησιμοποιηθούν στη μελέτη διάδοσης ενός σκουληκιού όταν κάποιοι υπολογιστές του δικτύου είναι προσωρινά σβηστοί αλλά δεν έχουν εγκαταστήσει κάποια επιδιόρθωση (patch) όπως για παράδειγμα στην περίπτωση του σκουληκιού Code Red I.

Έστω d ο μέσος βαθμός του δικτύου και γ ο ρυθμός με τον οποίο ένας μολυσματικός κόμβος αναρρώνει.Ο ρυθμός με τον οποίο προστίθενται καινούριοι μολυσμένοι κόμβοι στο δίκτυο είναι ανάλογος της εκτιμώμενης αναλογίας ευάλωτων κόμβων,του αριθμού μολυσμένων κόμβων και του ρυθμού μόλυνσης β.Από την άλλη πλευρά,ο ρυθμός με τον οποίο οι μολυσματικοί κόμβοι αναρρώνουν είναι ανάλογος με τον αριθμό των μολυσματικών κόμβων και το ρυθμό ανάρρωσης γ.

Βάσει όλων των προηγούμενων προκύπτει η παρακάτω διαφορική εξίσωση που περιγράφει το μοντέλο SIS (3):

$$\frac{di(t)}{dt} = \beta d(1-i(t))i(t) - \gamma i(t)$$

με αρχικές εξισώσεις

$$i(0) = \frac{I(0)}{N} > 0$$

και

$$\forall t \geq 0, i(t) + s(t) = 1$$

επίσης στη εξίσωση (3) έχουμε

$$\frac{di(t)}{dt} < 0 <=> s(t) < \frac{\gamma}{\beta d} = \delta$$

που σημαίνει ότι το σκουλήκι "πεθαίνει" αν η αρχική αναλογία ευάλωτων κόμβων s(t) είναι κάτω από το επιδημικό κατώφλι δ=γ/βd.

Η λύση της εξίσωσης (3) μας δίνει τον παρακάτω τύπο για τον υπολογισμό της αναλογίας των μολυσματικών κόμβων (έχει γίνει χρήση του τύπου του επιδημικού κατωφλίου δ).

$$i(t) = \frac{(1-\delta)i(0)}{i(0) + (1-\delta - i(0))e^{-(\beta' - \gamma)t}}$$

Στη περίπτωση που το σκουλήκι διαδίδεται σε πλήρη γράφο με N κόμβους όπου d=N-1 ,τότε το μοντέλο (3) γράφεται ως εξής(4):

$$\frac{di(t)}{dt} = \beta d(1-i(t))I(t) - \gamma i(t)$$

με λύση

$$i(t) = \frac{(1-\delta)i(0)}{i(0) + (1-\delta - i(0))e^{-[\beta(N-1)-\gamma]t}}$$

Ο Solomon[20]μελέτησε ένα διαφοροποιημένο μοντέλο με βαση το(4) όπου ο ρυθμός γ είναι ένας σταθμισμένος μέσος όρος του ρυθμού γ_1(για υπολογιστές που δεν έχουν αντιικό λογισμικό) εφαρμόσιμου στην αναλογία των μολυσματικών κόμβων και του ρυθμού γ_2 (για υπολογιστές με την πιο πρόσφατη έκδοση αντ-ικού λογισμικού) εφαρμόσιμου στην αναλογία των ευάλωτων κόμβων δηλαδή $\gamma = \gamma_1 i(t) + \gamma_2 (1-i(t))$.Με την συγκεκριμένη διαφοροποίηση βρέθηκε ότι η απαραίτητη αποτελεσματικότητα του αντι-ικού λογισμικού (που περιγράφεται από τον ρυθμο γ) πρέπει να είναι 0.5 για να σταματήσει η εξάπλωση πριν επιτευχθεί η εκθετική αύξηση.

Οι Kephart και White[21]χρησιμοποίησαν το μοντέλο (3) για την μελέτη της επιρροής τριών τοπολογιών στη διάδοση ιών:των ER τυχαίων γράφων,απλών πλεγμάτων με βαθμό 8 και ιεραρχικών τυχαίων γράφων.Οι σχετικές,με την παραπάνω μελέτη,λεπτομέρειες παραλείπονται μιάς και ξεφεύγουν από τα πλαίσια της συγκεκριμένης εργασίας.

Σύμφωνα με τους Wang και Wang[22],τα προηγούμενα μοντέλα ήταν περιορισμένης ακρίβειας λόγω της απλοϊκής αντιμετώπισης που είχαν απέναντι σε σημαντικούς χρονικούς παράγοντες όπως η καθυστέρηση μόλυνσης.Σαν καθυστέρηση μόλυνσης ορίζεται ο χρόνος από τη στιγμή που το σκουλήκι φτάνει στον κόμβο μέχρι τη στιγμή που ο κόμβος γίνεται μολυσματικός για τους γείτονες του.Οι Wang και Wang λοιπόν διαφοροποίησαν το μοντέλο (3) έτσι ώστε να συμπεριληφθεί και η καθυστέρηση μόλυνσης.Έτσι προέκυψε το μοντέλο (5):

$$\frac{di(t)}{dt} = \beta d e^{-\gamma\varepsilon}(1-i(t))i(t-\varepsilon) - \gamma i(t)$$

όπου για t<ε ισχύει i(t-ε)=0 και για t>=ε η αναλογία μολυσματικών κόμβων είναι ίδια με την αναλογία των μολυσματικών κόμβων για χρόνο t-ε , λόγω της καθυστέρησης ε.Ο όρος $e^{-\gamma}$ παριστάνει τη μετάβαση ενός μολυσματικού κόμβου στην ευάλωτη κατάσταση κατά τη διάρκεια της περιόδου καθυστέρησης. Η εξίσωση είναι μη γραμμική διαφορική εξίσωση και μπορεί να λυθεί κάτω από την υπόθεση ότι i(t-ε)=i(t) .

Η ομάδα που παρουσίασε το συγκεκριμένο μοντέλο,υποστήριξε την αναλυτική λύση με προσομοίωση και έδειξε ότι το επιδημικό κατώφλι δεν εξαρτάται μόνο από τον μέσο βαθμό του δικτύου αλλά και από την καθυστέρηση μόλυνσης.

Οι Pastor-Satorras και Vespignani[23]τροποποίησαν το μοντέλο (3),για την μελέτη της επιρροής της scale-free τοπολογίας Barabasi-Albert κατά τη διάρκεια της εξάπλωσης,με ρυθμό ανάρρωσης γ=1.Η scale-free κατανομή βαθμού δεν είναι συγκεντρωμένη γύρω από τη μέση τιμή της,γι'αυτό το λόγο το μοντέλο θα πρέπει να περιέχει διαφορικές εξισώσεις για κάθε ομάδα κόμβων βαθμού k.Το παρακάτω σύνολο διαφορικών εξισώσεων περιγράφει το τροποποιημένο μοντέλο(6):

$$\frac{di_k(t)}{dt} = \beta k(1-i_k(t))\Theta(\{i_k(t)\}_{d_{\min}}^{d_{\max}}) - i_k(t)$$

όπου η συνάρτηση

$$\Theta(\{i_k(t)\}_{d_{\min}}^{d_{\max}})$$

περιγράφει την πιθανότητα μία ακμή να είναι σύνδεση ενός κόμβου βαθμού k,και ισχύει:

$$\Theta(\{i_k(t)\}_{d_{\min}}^{d_{\max}}) = \frac{kP(k)}{d}$$

Συνεπώς η πιθανότητα μία ακμή να είναι σύνδεση ενός μολυσματικού κόμβου είναι

$$\Theta(t) = \frac{1}{d}\sum_{k=d_{\min}}^{d_{\max}} kP(k)i_k(t)$$

Η ομάδα της παραπάνω μελέτης κατέληξε ότι τα scale-free δίκτυα δεν έχουν επιδημικό κατώφλι.Η παρουσία όμως κρίσιμου σημείου στην scale-free κατανομή επιβάλλει την ύπαρξη μη μηδενικού κατωφλίου,γεγονός που προκάλεσε την διαφωνία μεταξύ των παραπάνω μελετητών.

Τα αποτελέσματα όλων των παραπάνω μελετών που βασίζονται στο μοντέλο SIS είναι άκρως πολύτιμα.Όμως λόγω της παρουσίας ανθρώπινων αντίμετρων,ένα μοντέλο στο οποίο οι κόμβοι που έχουν αναρρώσει δεν είναι πλέον ευάλωτοι,προσεγγίζει πιο πειστικά τη διαδικασία εξάπλωσης ενός σκουληκιού[16].

3.3 Μοντέλο SIR ή Kermack-McKendrick:

Σε αυτή τη κατηγορία μοντέλων ένας μολυσματικός κόμβος μπορεί να απομακρυνθεί.Αυτό το μοντέλο μπορεί να χρησιμοποιηθεί για την μελέτη της επιρροής της επιδιόρθωσης λογισμικού (software patching) και του μπλοκαρίσματος της κίνησης(traffic blocking).Σε οποιαδήποτε χρονική στιγμή ένας κόμβος μπορεί να

είναι ευάλωτος,μολυσματικός ή απομακρυσμένος.Έστω γ ο ρυθμός με τον οποίο οι μολυσματικοί κόμβοι απομακρύνονται.Τότε το μοντέλο SIR μπορεί να περιγραφεί από τις παρακάτω διαφορικές εξισώσεις(7):

$$\frac{di(t)}{dt} = \beta d(1-i(t))i(t) - \gamma i(t)$$
$$\frac{dr(t)}{dt} = \gamma i(t)$$

με αρχικές εξισώσεις

$$i(0) = \frac{I(0)}{N} \geq 0, r(0) = \frac{R(0)}{N} \geq 0$$

και

$$\forall t \geq 0 \; i(t) + s(t) = 1$$

Το επιδημικό κατώφλι των SIR μοντέλων είναι παρόμοιο με αυτό των SIS μοντέλων.

Οι Zou et al [24] χρησιμοποίησαν μία τροποποίηση του συστήματος εξισώσεων (7) για να προσδιορίσουν την επιρροή των ανθρώπινων αντίμετρων (μετακίνηση ευάλωτων και μολυσματικών κόμβων) και του πτωτικού ρυθμού β(t).Έτσι προέκυψε το two-factor μοντέλο που θα αναλυθεί λεπτομερώς παρακάτω.

Οι Boguna et al.[25]μελέτησαν το SIR μοντέλο με πιθανότητα γ=1 σε scale-free δίκτυα.Χρησιμοποιώντας τους συμβολισμούς της υποενότητας SIS παραπάνω,το μοντέλο μπορεί να περιγραφεί με τον εξής τρόπο(8):

$$\frac{di_k(t)}{dt} = \beta k(1-i_k(t))\Theta(\{i_k(t)\}_{d_{\min}}^{d_{\max}}) - i_k(t)$$
$$\frac{dr_k(t)}{dt} = \gamma i_k(t)$$

που μπορεί να λυθεί αν θεωρηθεί ότι το i(0) είναι πολύ μικρό στην αρχή της διαδικασίας εξάπλωσης του σκουληκιού.

Οι Pastor-Satorras και Vespignani[23]έκαναν μία μελέτη-προσομοίωση για

την εξέταση της επιρροής της ανοσοποίησης των κόμβων (δηλαδή απομακρυνση κόμβων) κατά τη διαδικασία διάδοσης του σκουληκιού,πριν το σκουλήκι εμφανιστεί στο δίκτυο.Έδειξαν ότι η τυχαία ανοσοποίηση είναι ανίκανη να καθυστερήσει τη διάδοση.Παρόλα αυτά,η ανοσοποίηση των κόμβων που έχουν τους μεγαλύτερους βαθμούς,έχει τη δυνατότητα να εμποδίσει την ανάπτυξη της διάδοσης.Αν και αυτό το αποτέλεσμα φαίνεται ενδιαφέρον,οι προαναφερθέντες ερευνητές δηλώνουν ότι η ανίχνευση κόμβων με μεγάλο βαθμό,στα scale-free δίκτυα είναι δύσκολη υπόθεση.Κύριο μειονέκτημα της μελέτης των Pastor-Satorras και Vespignani είναι ότι η ανοσοποίηση θεωρείται στατική δηλαδή,ένα ποσοστό κόμβων ανοσοποιείται πριν ξεκινήσει η διαδικασία εξάπλωσης του σκουληκιού.Στην πραγματικότητα τα αντίμετρα αυτά πρέπει να είναι δυναμικά ώστε να είναι ικανά να επιβραδύνουν την εξάπλωση του σκουληκιού[16].

3.4 Μοντέλο SIDR:

Το συγκεκριμένο μοντέλο αναλύθηκε από τους William και Leveille[26]με σκοπό τον υπολογισμό της αποτελεσματικότητας ενός μηχανισμού που ονομάζεται virus throttling.Το virus throttling είναι ένας αυτόματος μηχανισμός για της επιβράδυνση της εξάπλωσης του σκουληκιού.Στο συγκεκριμένο μοντέλο ένας κόμβος μπορεί να βρίσκεται σε μία από τις τέσσερις καταστάσεις: ευάλωτος,μολυσματικός,ανιχνευμένος(όπου το σκουλήκι έχει ανιχνευθεί και δεν μπορεί να εξαπλωθεί παραπάνω) και απομακρυσμένος.Το μοντέλο θεωρεί ότι το δίκτυο που γίνεται η εξάπλωση είναι ένας πλήρης γράφος.Επιπλέον διακρίνονται δύο φάσεις στο SIDR μοντέλο.Κατα την πρώτη φάση,πριν την απελευθέρωση της υπογραφής του σκουληκιού,οι κόμβοι μεταβαίνουν με ρυθμό β από την κατάσταση που είναι ευάλωτοι στην κατάσταση που είναι μολυσματικοί.Στην δεύτερη φάση,μετά από κάποιο χρόνο από την έναρξη της εξάπλωσης ,το σκουλήκι ανιχνεύεται με κάποιο ρυθμό γ.Η μελέτη επικεντρώνεται σε δύο ποσότητες.Στον αριθμό των μολυσματικών κόμβων και στη διάρκεια της εξάπλωσης.Στο συγκεκριμένο μοντέλο οι κόμβοι χωρίζονται σε δύο κατηγορίες throttled και unthrottled.Αν ένας throttled κόμβος μολυνθεί,δεν μπορεί να διαδόσει την μόλυνση και μεταβαίνει απ'ευθείας στην κατάσταση ανιχνευμένος.Τα αποτελέσματα της συγκεκριμένης μελέτης έδειξαν ότι

όταν παραπάνω από τους μισούς κόμβους είναι throttled,ακόμα και μία καθυστερημένη απελευθέρωση της υπογραφής του σκουληκιού θα προκαλέσει μικρό ξέσπασμα της μόλυνσης.

3.5 Μοντέλο SIRS:

Οι Wang και Wang[22]χρησιμοποίησαν μία τροποποιημένη μορφή του μοντέλου SIS ώστε να μελετήσουν την ετοιμότητα του κόμβου απέναντι στην μόλυνση.Από τη στιγμή που ένας μολυσματικός κόμβος απομακρυνθεί,παραμένει σε αυτή την κατάσταση για χρόνο v (περίοδος ετοιμότητας ή επαγρύπνησης) και μετά το πέρας της περιόδου αυτής ο απομακρυσμένος κόμβος γίνεται ευάλωτος ξανά.Στο συγκεκριμένο μοντέλο η ευπάθεια ενός κόμβου μοντελοποιείται μέσω μιας παραμέτρου Φ που παίρνει τιμές από 0 (για παντελή ευπάθεια) έως 1 (για ανοσία).Το μοντέλο αυτό περιγράφεται από την παρακάτω μη γραμμική διαφορική εξίσωση:

$$\frac{di(t)}{dt} = \beta d(1 - i(t) - \int_{t-v}^{t} i(t))i(t) - \gamma i(t)$$

της οποίας η λύση δείχνει ότι ο αριθμός των μολυσματικών κόμβων μειώνεται όσο η περίοδος επαγρύπνησης αυξάνεται.Αξίζει να αναφερθεί,ότι η περίοδος επαγρύπνησης δεν έχει καμία επίδραση στο επιδημικό κατώφλι.

3.6 Μοντέλο RCS (Random Constant Spread):

Όπως αναφέραμε στην ανάλυση του μοντέλου SI,οι Staniford,Paxson και Weaver[17] χρησιμοποίησαν μία τροποποιημένη μορφή αυτού για την μελέτη εξάπλωσης του Code Red σκουληκιού.Με αυτόν τον τρόπο αναπτύχθηκε το μοντέλο RCS χρησιμοποιώντας εμπειρικά δεδομένα από την έξαρση του σκουληκιού Code Red.Η ανάπτυξη του μοντέλου RCS βασίστηκε στην δεύτερη έκδοση του σκουληκιού Code Red,στην οποία είχε διορθωθεί μία ατέλεια που υπήρχε στην πρώτη έκδοση αυτού.Η ατέλεια σχετιζόταν με το γεγονός ότι η γεννήτρια αριθμών (που χρησίμευε στην ανίχνευση IP διευθύνσεων ευάλωτων τερματικών) παρήγαγε ακριβώς την ίδια ακολουθία αριθμών (δηλαδή IP διευθύνσεων) σε κάθε αντίγραφο του σκουληκιού.Αυτό είχε σαν αποτέλεσμα την γραμμική εξάπλωση του σκουληκιού και

συνεπώς την αδυναμία του να μολύνει πολλά τερματικά.

Στο μοντέλο RCS γίνεται η υπόθεση ότι το σκουλήκι είναι εφοδιασμένο με μία καλή (απουσία του bug της πρώτης έκδοσης) γεννήτρια τυχαίων αριθμών και με N ορίζεται το συνολικό πλήθος ευάλωτων συστημάτων που μπορούν να μολυνθούν μέσω Internet.

Το πλήθος N θεωρείται σταθερό,αγνοώντας την πιθανή επιδιόρθωση των συστημάτων κατά τη διαδικασία εξάπλωσης του σκουληκιού,την απομάκρυνση αυτών καθώς και την πιθανή απενεργοποίηση τους κατά τη διάρκεια της νύχτας.Είναι προφανές λοιπόν ότι το συγκεκριμένο μοντέλο δεν λαμβάνει καθόλου υπόψιν τα πιθανά αντίμετρα,από πλευράς λογισμικού και από πλευράς χρήστη, στην εξάπλωση του σκουληκιού.

Με K υποδηλώνεται ο αρχικός ρυθμός μόλυνσης των ευάλωτων κόμβων και λαμβάνεται σαν σταθερά,χωρίς να βασίζεται σε επεξεργαστική ισχύ,σύνδεση δικτύου ή τοποθεσία της μολυσματικής συσκευής.Η ανίχνευση ευάλωτων συσκευών για μόλυνση γίνεται τελείως στην τύχη και από τη στιγμή που μολυνθεί κάποια από αυτές,παραμένει σε αυτή την κατάσταση για πάντα.

Με T υποδηλώνεται ο χρόνος στον οποίο η μόλυνση φτάνει το μέγιστο σημείο της εξάπλωσής της,με α το ποσοστό των ευάλωτων συσκευών που έχουν μολυνθεί και με t ο χρόνος.

Οι Staniford et al. κατέληξαν στο συμπέρασμα ότι η σχέση

$$Nda = (Na)K(1-a)dt$$

δείχνει πόσες παραπάνω συσκευές (Ndα) θα μολυνθούν στην επόμενη ποσότητα χρόνου (dt) σε συγκεκριμένο χρόνο t δεδομένου ότι έχουν μολυνθεί α% των συνολικών ευάλωτων συσκευών.

Η παραπάνω σχέση δίνει την διαφορική εξίσωση

$$\frac{da}{dt} = aK(1-a)$$

με λύση

$$a = \frac{e^{K(t-T)}}{1+e^{K(t-T)}}$$

Το μοντέλο αυτό μπορεί να προβλέψει τον αριθμό των μολυσματικών συσκευών σε χρόνο t όταν το K είναι γνωστό.Όσο μεγαλύτερο είναι το K,τόσο πιο γρήγορα το σκουλήκι θα φτάσει στο σημείο ισορροπίας.

3.7 Μοντέλο AAWP (Analytical Active Worm Propagation):

Το μοντέλο AAWP είναι ένα ντετερμινιστικό μοντέλο διακριτού χρόνου που παρουσιάστηκε από τους Chen,Gao και Kwiat[27]και περιγράφει την εξάπλωση των σκουληκιών που χρησιμοποιούν τυχαία σάρωση.Μία τροποποίηση του συγκεκριμένου μοντέλου είναι το Local AAWP το οποίο αφορά σε σκουλήκια που χρησιμοποιούν μηχανισμό σάρωσης τοπικού υποδικτύου.

Για το μοντέλο AAWP οι Chen,Gao και Kwiat θεώρησαν τα εξής χαρακτηριστικά για το σκουλήκι.Αρχικά το σκουλήκι έχει ήδη μολύνει όλες τις συσκευές που βρίσκονται σε μία hitlist και έπειτα χρησιμοποιείται μηχανισμός τυχαίας σάρωσης για ανίχνευση και μόλυνση άλλων ευάλωτων συσκευών.Με άλλα λόγια ο χρόνος που χρειάζεται το σκουλήκι για τη μόλυνση της hitlist δεν λαμβάνεται υπόψιν γιατί θεωρείται αμελητέος,μιας και η hitlist αποκτάται απευθείας με την απελευθέρωση του σκουληκιού και η μόλυνση επιτυγχάνεται σε πολύ μικρό χρονικό διάστημα.Επιπλέον τα σκουλήκια έχουν τη δυνατότητα σάρωσης πολλών συσκευών ταυτόχρονα και δεν ξαναμολύνουν αυτές που έχουν ήδη μολυνθεί.Τέλος το οποιοδήποτε σκουλήκι,ολοκληρώνει την επικοινωνία του με την συσκευή σε μία μονάδα χρόνου.Πιο συγκεκριμένα όταν γίνεται σάρωση και ανιχνεύεται μία συσκευή,ανεξαρτήτως αν αυτή είναι ευάλωτη,μολυσματική ή με αχρησιμοποίητη IP διεύθυνση,το σκουλήκι έχει τελειώσει την οποιαδήποτε διαδικασία μαζί της σε μία μόνο μονάδα χρόνου.Λόγω τυχαίας σάρωσης,η πιθανότητα να ανιχνευθεί ένας υπολογιστής είναι $1/2^{32}$ (IPv4).Με m_i παριστάνεται ο αριθμός των ευάλωτων υπολογιστών(μαζί με τους μολυσματικούς) σε χρόνο i>=0 και με ni ο αριθμός των

μολυσματικών υπολογιστών σε χρόνο i.Πριν την εξάπλωση του σκουληκιού (i=0) ισχύει m_0=N και n_0=h (όπου N είναι το πλήθος των των ευάλωτων υπολογιστών και h το μέγεθος της hitlist).

Σύμφωνα με την ομάδα παρουσίασης του συγκεκριμένου μοντέλου,αν υπάρχουν mi ευάλωτοι υπολογιστές (μαζί με τους μολυσματικούς) και n_i μολυσματικοί υπολογιστές,τότε κατά μέσο όρο την επόμενη χρονική στιγμή θα υπάρχουν στο δίκτυο

$$(m_i - n_i)[1-(1-\frac{1}{2^{32}})^{sn_i}]$$

υπολογιστές που μόλις μολύνθηκαν,όπου s είναι ο ρυθμός σάρωσης.

Δεδομένου ρυθμού θανάτου d (ο ρυθμός με τον οποίο μία μόλυνση ανιχνεύεται σε έναν υπολογιστή και εξουδετερώνεται χωρίς χρήση επιδιόρθωσης) και ρυθμού επιδιόρθωσης p,την επόμενη χρονική στιγμή θα υπάρχουν dn_i+pn_i μολυσματικοί υπολογιστές όπου άλλοι θα γίνουν ευάλωτοι και άλλοι μη ευάλωτοι και συνεπώς το συνολικό πλήθος των ευάλωτων υπολογιστών (μαζί με τους μολυσματικούς) θα γίνει (1-p)m_i.Άρα την επόμενη χρονική στιγμή το πλήθος των μολυσματικών υπολογιστών θα είναι

$$n_{i+1} = n_i + (m_i - n_i)[1-(1-\frac{1}{2^{32}})^{sn_i}] - (d+p)n_i$$

την ίδια στιγμή που

$$m_{i+1} = (1-p)m_i <=> m_i = (1-p)^i m_0 = (1-p)^i N$$

και τελικά

$$n_{i+1} = (1-d-p)n_i + [(1-p)^i N - n_i][1-(1-\frac{1}{2^{32}})^{sn_i}]$$

με i>=0 και n_0=h.

Οι συγγραφείς κατέληξαν στο συμπέρασμα ότι το μοντέλο AAWP είναι πιο ακριβές από τα επιδημιολογικά μοντέλα.Το συμπέρασμα αυτό βασίζεται στο γεγονός

ότι το συγκεκριμένο μοντέλο είναι διακριτού χρόνου,λαμβάνει υπόψιν του τον χρόνο που χρειάζεται ένα σκουλήκι για να μολύνει έναν υπολογιστή όπως επίσης και τον ρυθμό επιδιόρθωσης αυτού και τέλος θεωρεί ότι ένας ευάλωτος υπολογιστής μπορεί να ανιχνευθεί από δύο ή και περισσότερους μολυσματικούς υπολογιστές την ίδια χρονική στιγμή.

3.8 Μοντέλο Two factor (Two factor worm model):

Οι Zou,Gong και Towsley[24]με αφορμή την έξαρση του Code Red σκουληκιού, παρουσίασαν ένα μοντέλο εξάπλωσης σκουληκιών που μπορεί να χρησιμοποιηθεί χωρίς τοπολογικούς περιορισμούς ,το επονομαζόμενο Two factor worm model.Το συγκεκριμένο μοντέλο που αποτελεί μία τροποποίηση του επιδημιολογικού μοντέλου SIR λαμβάνει υπόψιν του δύο παράγοντες,που σύμφωνα με την παραπάνω ερευνητική ομάδα, είναι ύψιστης σημασίας για τη μοντελοποίηση ενός σκουληκιού σε ένα δίκτυο.

α)Τα ανθρώπινα αντίμετρα σχετίζονται με την απομάκρυνση τερματικών από το δίκτυο,τα οποία είναι είτε ευάλωτα,είτε μολυσματικά.Πιο συγκεκριμένα τα ανθρώπινα αντίμετρα είναι ενέργειες στις οποίες καταφεύγουν οι χρήστες,όταν κατά την διάρκεια εξάπλωσης ενός σκουληκιού,το ευρύ κοινό μαθαίνει για αυτό.Παραδείγματα ανθρώπινων αντιμέτρων είναι ο καθαρισμός μολυσμένων τερματικών,η αναβάθμιση ή η επιδιόρθωση τρωτών σημείων ενός ευάλωτου τερματικού,το μπλοκάρισμα της κίνησης του σκουληκιού μέσω τοίχων προστασίας ή δρομολογητών,ακόμα και η αποσύνδεση των τερματικών από το δίκτυο.

β)Ο ρυθμός μόλυνσης β(t) κατά τη διάρκεια εξάπλωσης ενός σκουληκιού δεν είναι σταθερός.Όσο περνάει ο χρόνος από την απελευθέρωση του σκουληκιού,τόσο ο ρυθμός μόλυνσης μειώνεται.Αυτό δικαιολογείται από το γεγονός ότι ,κατά τη διάδοση ενός σκουληκιού σε ένα δίκτυο μεγάλης κλίμακας,η κίνηση αυξάνεται κατά πολύ με αποτέλεσμα την υπερφόρτωσή του.

Με βάση τα προαναφερθέντα,ο ρυθμός β μοντελοποιείται συναρτήσει του

χρόνου t και η απομάκρυνση των κόμβων λόγω ανθρώπινων αντιμετρων χωρίζεται σε απομάκρυνση μολυσματικών κόμβων R(t) και σε απομάκρυνση ευάλωτων κόμβων Q(t).

Συμφωνα με το SIR μοντέλο (στο οποίο βασίζεται) ,η αλλαγή του αριθμού των ευάλωτων κόμβων από χρόνο t μέχρι t+Δt περιγράφεται από την εξίσωση

$$S(t+\Delta t)-S(t)=-\beta(t)S(t)I(t)\Delta t-\frac{dQ(t)}{dt}\Delta t$$

και συνεπώς

$$\frac{dS(t)}{dt}=-\beta(t)S(t)I(t)-\frac{dQ(t)}{dt}$$

Δεδομένου ότι S(t)+I(t)+R(t)+Q(t)=N <=> S(t)=N-I(t)-R(t)-Q(t) η παραπάνω διαφορική εξίσωση δίνει την εξίσωση που περιγράφει τη συμπεριφορά του αριθμού των μολυσματικών κόμβων I(t):

$$\frac{dI(t)}{dt}=\beta(t)[N-R(t)-I(t)-Q(t)]I(t)-\frac{dR(t)}{dt}$$

η οποία περιγράφει το μοντέλο two factor.

Για την επίλυση της παραπάνω διαφορικής εξίσωσης,πρέπει να καθοριστούν οι τύποι υπολογισμού των β(t),R(t) και Q(t).

Το β(t) καθορίζεται από την επιρροή της κίνησης,που προκαλεί το σκουλήκι,στο δίκτυο και από την αποτελεσματικότητα εξάπλωσης του κακόβουλου κώδικά του.

Τα R(t) και Q(t) σχετίζονται με την αντίληψη των ανθρώπων για το σκουλήκι καθώς και με τις δυσκολίες επιδιόρθωσης και φιλτραρίσματος.

Για τη διαδικασία απομάκρυνσης μολυσματικών κόμβων γίνεται χρήση των υποθέσεων του SIR μοντέλου και συνεπώς προκύπτει ότι

$$\frac{dR(t)}{dt} = \gamma I(t)$$

Για τη διαδικασία απομάκρυνσης ευάλωτων κόμβων πρέπει να ληφθεί υπόψιν ότι στην αρχή οι άνθρωποι έχουν πλήρη άγνοια για το σκουλήκι,που σημαίνει ότι ο αριθμός των ευάλωτων κόμβων που απομακρύνονται στην αρχή είναι πολύ μικρός και αυξάνεται με αργό ρυθμό.Όσο ο χρόνος περνάει και το σκουλήκι αρχίζει να γίνεται γνωστό (με τη μόλυνση όλο και περισσότερων κόμβων) ,η ταχύτητα ανοσοποίησης αυξάνεται.Προς το τέλος της διαδικασίας εξάπλωσης του σκουληκιού,η ταχύτητα ανοσοποίησης μειώνεται,μιας και ο αριθμός των ευάλωτων κόμβων μικραίνει,και τείνει στο μηδέν όταν πλέον δεν υπάρχουν καθόλου ευάλωτοι κόμβοι.Βάσει της παραπάνω περιγραφής,η συμπεριφορά του Q(t),μοιάζει με μία τυπική επιδημική εξάπλωση.Γι'αυτό τον λόγο,για την μοντελοποίηση του Q(t) χρησιμοποιείται το μοντέλο SI και συνεπώς προκύπτει

$$\frac{dQ(t)}{dt} = \mu S(t) J(t)$$

όπου J(t)=I(t)+R(t).

Τέλος για τον πτωτικό ρυθμό μόλυνσης β(t) χρησιμοποιείται η εξίσωση

$$\beta(t) = \beta_0 [1 - \frac{I(t)}{N}]^n$$

όπου β_0 είναι ο αρχικός ρυθμός μόλυνσης και ο εκθέτης n έχει χρησιμοποιηθεί για την προσαρμογή της ευαισθησίας του ρυθμού μόλυνσης στον αριθμό των μολυσματικών κόμβων I(t).Για n=0 έχουμε σταθερό ρυθμό μόλυνσης.

Συνεπώς το σύνολο των διαφορικών εξισώσεων που περιγράφουν το μοντέλο two factor είναι

$$\frac{dS(t)}{dt} = -\beta(t)S(t)I(t) - \frac{dQ(t)}{dt}$$
$$\frac{dR(t)}{dt} = \gamma I(t)$$
$$\frac{dQ(t)}{dt} = \mu S(t)J(t)$$
$$\beta(t) = \beta_0[1 - \frac{I(t)}{N}]^n$$
$$N = S(t) + I(t) + R(t) + Q(t)$$
$$I(0) = I_0 << N \ , \ S(0) = N - I_0 \ , \ R(0) = Q(0) = 0$$

Στην παρουσίαση του συγκεκριμένου μοντέλου η εξάπλωση ενός σκουληκιού,παρόλο που αντιμετωπίζεται συνήθως σαν διακριτού χρόνου διαδικασία,προσεγγίζεται σαν συνεχής διαδικασία.Μία τέτοια προσέγγιση είναι ακριβής σε δίκτυα μεγάλης κλίμακας και χρησιμοποιείται κατά κόρον στην επιδημιολογική μοντελοποίηση.

3.9 Μοντέλο compartment-based:

Το μοντέλο compartment based παρουσιάστηκε από τους Serazzi και Zanero[28]κάτω από την ιδέα ότι ολόκληρο το Internet είναι αδύνατον να μοντελοποιηθεί σαν γράφος όπου κάθε κόμβος του αναπαριστά ένα τερματικό ή έναν δρομολογητή.Με βάση την παραπάνω ιδέα,το Internet προσεγγίστηκε μακροσκοπικά σαν διασύνδεση ενός αριθμού αυτόνομων συστημάτων (Autonomous systems-AS),τα οποία ουσιαστικά είναι υποδίκτυα που διαχειρίζονται από μία και μόνο αρχή(authority).

Στο μοντέλο αυτό, η εξάπλωση του σκουληκιού εξετάζεται

α) στο εσωτερικό ενός αυτόνομου συστήματος (ή στο εσωτερικό μίας περιοχής εντός του αυτόνομου συστήματος,η οποία χαρακτηρίζεται από πυκνές συνδέσεις) όπου το σκουλήκι διαδίδεται ανεμπόδιστο και η διάδοση του ακολουθεί το μοντέλο RCS και β) ανάμεσα στα αυτόνομα συστήματα όπου χρειάζεται η επέκταση της εξίσωσης dα/dt=Kα(1-α) του μοντέλου RCS.

Έστω N_i ο αριθμός των ευάλωτων τερματικών στο AS_i ,α_i το ποσοστό των

μολυσματικών τερματικών στο AS_i και K η μέση ταχύτητα διάδοσης του σκουληκιού που σε πρώτη φάση θεωρείται σταθερή σε κάθε AS.

Με $P_{IN,i}$ συμβολίζεται η πιθανότητα ένα τερματικό να επιτεθεί σε ένα άλλο εντός του ίδιου AS και με $P_{OUT,i}$ η πιθανότητα να επιτεθεί σε ένα τερματικό άλλου AS.

Σε ένα απλό μοντέλο με μόλις δύο αυτόνομα συστήματα,η ακόλουθη εξίσωση περιγράφει τις εσωτερικές και εξωτερικές προσπάθειες του σκουληκιού για μόλυνση του AS_1

$$N_1 da_1 = [N_1 a_1 K P_{IN,1} dt + N_2 a_2 K P_{OUT,2} dt](1-a_1)$$

ομοίως περιγράφονται και οι προσπάθειες μόλυνσης προς το AS_2 και συνεπώς προκύπτει το παρακάτω σύστημα δύο εξισώσεων

$$\frac{da_1}{dt} = [a_1 K P_{IN,1} + \frac{N_2}{N_1} a_2 K P_{OUT,2}](1-a_1)$$
$$\frac{da_2}{dt} = [a_2 K P_{IN,2} + \frac{N_1}{N_2} a_1 K P_{OUT,1}](1-a_2)$$

Θεωρώντας ότι το σκουλήκι γεννά τυχαία την διεύθυνση IP του στόχου έχουμε ότι $P_{IN,1}=N_1/N$ και $P_{OUT,1}=1-P_{IN,1}=N_2/N$.

Βάσει των παραπάνω ισοτήτων και επεκτείνοντας το σύστημα εξισώσεων για i=1,...n έχουμε ότι

$$\frac{da_i}{dt} = [a_i K \frac{N_i}{N} + \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{N_j}{N_i} a_j K \frac{N_i}{N}](1-a_i)$$

Το αποτέλεσμα ολοκλήρωσης κάθε μίας από τις παραπάνω εξισώσεις είναι μία λογιστική συνάρτηση (παρόμοια αυτής που προκύπτει από το μοντάλο RCS) που ωθείται στην αύξηση της από τον δεύτερο προσθεταίο της παράστασης που αντιπροσωπεύει τις εισερχόμενες επιθέσεις που προέρχονται έξω από το αυτόνομο σύστημα.Απλοποιώντας την παραπάνω εξίσωση προκύπτει

$$\frac{da_i}{dt} = [a_i K \frac{N_i}{N} + \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{N_i}{N} a_j K](1-a_i)$$

όπου παραλείφθηκε ο όρος που περιγράφει τον ρυθμό εισερχόμενων επιθέσεων και με ακόμα μεγαλύτερη απλοποίηση προκύπτει

$$\frac{da_i}{dt} = [\sum_{\substack{j=1 \\ j\neq i}}^{n} N_j a_j](1-a_i)\frac{K}{N}$$

που είναι ένα μη γραμμικό σύστημα διαφορικών εξισώσεων του οποίου τα αποτελέσματα αποτελούν μία λύση για το μοντέλο με παραμέτρους K και N.

Θεωρώντας,τέλος ότι

$a = \frac{\sum_{i=1}^{n} N_i a_i}{N}$ έχουμε ότι $\frac{da}{dt} = \frac{d}{dt}[\frac{\sum_{i=1}^{n} N_i a_i}{N}] = \frac{1}{N}\sum_{i=1}^{n} N_i \frac{da_i}{dt}$

και με χρήση της τελευταίας εξίσωσης προκύπτει η εξίσωση του RCS μοντέλου dα/dt=Kα(1-α).

3.10 Σύνοψη:

<u>SI μοντελο:</u>

- Ένας κόμβος μπορεί να βρίσκεται σε μία εκ τις δύο καταστάσεις:ευάλωτος(S)-μολυσματικός(I)
- Από τη στιγμή που ένας ευάλωτος κόμβος γίνει μολυσματικός,η κατάσταση του παραμένει μολυσματική καθόλη τη διάρκεια της εξάπλωσης του σκουληκιού
- Μελέτη της χειρότερης περίπτωσης μίας διάδοσης,όταν δεν υπάρχουν αυτοματοποιημένα και ανθρώπινα αντίμετρα.
- Το SI μοντέλο περιγράφεται από τη διαφορική εξίσωση

$$\frac{di(t)}{dt} = \beta d(1-i(t))i(t)$$

- Η λύση είναι η λογιστική καμπύλη(η πιο κοινή σιγμοειδής καμπύλη)

$$i(t) = \frac{i(0)e^{\beta' t}}{1 - i(0) + i(0)e^{\beta' t}}$$

όπου β' = βd.

- Η παραπάνω σιγμοειδής καμπύλη μπορεί να χωριστεί σε τρεις περιοχές[16]:

  α) αργή εκκίνηση,όπου ελάχιστοι κόμβοι μολύνονται σε κάθε βήμα χρόνου.
  β) εκθετική αύξηση,όπου ο αριθμός των κόμβων που μόλις μολύνθηκαν αυξάνεται εκθετικά.
  γ) κατάσταση ισορροπίας,όπου ο αριθμός των μολυσματικών κόμβων παίρνει κάποια μέγιστη τιμή γύρω από την οποία ταλαντώνεται σταθερά

SIS μοντέλο:

- Σε αυτή τη κατηγορία μοντέλων,ένας μολυσματικός κόμβος αναρρώνει και συνεπώς γίνεται ευάλωτος ξανά.Αυτά τα μοντέλα μπορούν να χρησιμοποιηθούν στη μελέτη διάδοσης ενός σκουληκιού όταν κάποιοι υπολογιστές του δικτύου είναι προσωρινά σβηστοί αλλά δεν έχουν εγκαταστήσει κάποια επιδιόρθωση (patch)(περίπτωση Code Red I)

$$\frac{di(t)}{dt} = \beta d(1 - i(t))i(t) - \gamma i(t)$$

$$\frac{di(t)}{dt} < 0 <=> s(t) < \frac{\gamma}{\beta d} = \delta$$

  που σημαίνει ότι το σκουλήκι "πεθαίνει" αν η αρχική αναλογία ευάλωτων κόμβων s(t) είναι κάτω από το επιδημικό κατώφλι δ.
- Λόγω της παρουσίας ανθρώπινων αντίμετρων,ένα μοντέλο στο οποίο οι κόμβοι που έχουν αναρρώσει δεν είναι πλέον ευάλωτοι,προσεγγίζει πιο πειστικά τη διαδικασία εξάπλωσης ενός σκουληκιού[16].

SIR (Kermack-McKendrick) μοντέλο:

- Ευάλωτος,μολυσματικός ή απομακρυσμένος.
- Χρησιμοποιείται για την μελέτη της επιρροής του software patching και του traffic blocking στην εξάπλωση του σκουληκιού.

- $$\frac{di(t)}{dt} = \beta d(1-i(t))i(t) - \gamma i(t)$$
  $$\frac{dr(t)}{dt} = \gamma i(t)$$
- Το επιδημικό κατώφλι των SIR μοντέλων είναι παρόμοιο με αυτό των SIS μοντέλων.

SIDR μοντέλο:

- William και Leveille[26].
- Σκοπός:Ο υπολογισμός της αποτελεσματικότητας ενός μηχανισμού που ονομάζεται virus throttling(αυτόματος μηχανισμός για την επιβράδυνση της εξάπλωσης του σκουληκιού)
- Ευάλωτος,μολυσματικός,ανιχνευμένος και απομακρυσμένος.
- Τοπολογία δικτύου εξάπλωσης:πλήρης γράφος.
- Πρώτη φάση,πριν γίνει γνωστο το σκουλήκι,οι κόμβοι μεταβαίνουν με ρυθμό β από την κατάσταση που είναι ευάλωτοι στην κατάσταση που είναι μολυσματικοί.
- Δεύτερη φάση,μετά από κάποιο χρόνο από την έναρξη της εξάπλωσης ,το σκουλήκι ανιχνεύεται με κάποιο ρυθμό γ.
- Δύο είδη κόμβων:throttled-αν μολυνθούν,δεν διαδίδουν τη μόλυνση και μεταβαίνουν στην κατάσταση ‘‘ανιχνευμένος’’ και unthrottled.
- Αν πάνω από τους μισούς κόμβους είναι throttled και η απελευθέρωση υπογραφής του σκουληκιού γίνει καθυστερημένα,τότε το ξέσπασμα της μόλυνσης θα είναι πολύ μικρό.

SIRS μοντέλο:

- Wang και Wang[22]-τροποποίηση SIS μοντέλου.
- Σκοπός: Μελέτη ετοιμότητας του κόμβου απέναντι στην μόλυνση.
- Ευάλωτος,μολυσματικός,απομακρυσμένος,ευάλωτος.

- Όταν ένας μολυσματικός κόμβος απομακρυνθεί,παραμένει σε αυτή την κατάσταση για χρόνο v (περίοδος ετοιμότητας ή επαγρύπνησης) και μετά το πέρας της περιόδου αυτής ο απομακρυσμένος κόμβος γίνεται ευάλωτος ξανά.
- Μοντελοποίηση ευπάθειας μέσω παραμέτρου Φ με τιμές από 0 (για παντελή ευπάθεια) έως 1 (για ανοσία)

$$\frac{di(t)}{dt} = \beta d(1 - i(t) - \int_{t-v}^{t} i(t))i(t) - \gamma i(t)$$

- Ο αριθμός των μολυσματικών κόμβων μειώνεται όσο η περίοδος επαγρύπνησης αυξάνεται.

RCS μοντέλο:

- Staniford,Paxson και Weaver[17]-τροποποίηση SI μοντέλου.
- Σκοπός: Μελέτη εξάπλωσης Code Red v.2 σκουληκιού

$$\frac{da}{dt} = aK(1 - a)$$

- Όσο μεγαλύτερο είναι το K,τόσο πιο γρήγορα το σκουλήκι θα φτάσει στο σημείο ισορροπίας.
- Υπόθεση ότι το σκουλήκι είναι εφοδιασμένο με μία καλή (απουσία του bug της πρώτης έκδοσης- παραγωγή ακριβώς ίδιας ακολουθίας αριθμών (δηλαδή IP διευθύνσεων) σε κάθε αντίγραφο του σκουληκιού) γεννήτρια τυχαίων αριθμών.
- Το συγκεκριμένο μοντέλο δεν λαμβάνει καθόλου υπόψιν τα πιθανά αντίμετρα,από πλευράς λογισμικού και από πλευράς χρήστη, στην εξάπλωση του σκουληκιού.

AAWP (Analytical Active Worm Propagation) μοντέλο:

- Chen,Gao και Kwiat [27]
- Διακριτού χρόνου.
- Σκοπός: Μελέτη εξάπλωσης σκουληκιών που χρησιμοποιούν τυχαία σάρωση.
- Υπόθεση ότι το σκουλήκι έχει ήδη μολύνει όλες τις συσκευές που βρίσκονται σε μία hitlist και έπειτα χρησιμοποιείται μηχανισμός τυχαίας σάρωσης για ανίχνευση και μόλυνση άλλων ευάλωτων συσκευών.Επιπλέον τα σκουλήκια έχουν τη δυνατότητα σάρωσης πολλών συσκευών ταυτόχρονα και δεν

ξαναμολύνουν αυτές που έχουν ήδη μολυνθεί.Τέλος το οποιοδήποτε σκουλήκι,ολοκληρώνει την επικοινωνία του με την συσκευή σε μία μονάδα χρόνου.

- Δεδομένου ρυθμού θανάτου d και ρυθμού επιδιόρθωσης p,ο αριθμός των μολυσματικών κόμβων στο δίκτυο την επόμενη χρονική στιγμή δίνεται από τον τύπο

$$n_{i+1} = (1-d-p)n_i + [(1-p)^i N - n_i][1-(1-\frac{1}{2^{32}})^{sn_i}]$$

Με n_0=h και i>=0.

- Το συγκεκριμένο μοντέλο είναι διακριτού χρόνου,λαμβάνει υπόψιν του τον χρόνο που χρειάζεται ένα σκουλήκι για να μολύνει έναν υπολογιστή όπως επίσης και τον ρυθμό επιδιόρθωσης αυτού και τέλος θεωρεί ότι ένας ευάλωτος υπολογιστής μπορεί να ανιχνευθεί από δύο ή και περισσότερους μολυσματικούς υπολογιστές την ίδια χρονική στιγμή.Άρα η ακρίβειά του είναι μεγαλύτερη από αυτή των επιδημιολογικών μοντέλων.

Μοντέλο Two-factor:

- Zou,Gong και Towsley[24]-τροποποίηση του SIR
- Αφορμή η έξαρση του Code Red σκουληκιού.Είναι ένα μοντέλο εξάπλωσης σκουληκιών που μπορεί να χρησιμοποιηθεί χωρίς τοπολογικούς περιορισμούς.
- Υποθέσεις: α)Τα ανθρώπινα αντίμετρα που σχετίζονται με την απομάκρυνση τερματικών από το δίκτυο,είτε ευάλωτων(Q(t)),είτε μολυσματικών(R(t)).Για παράδειγμα ο καθαρισμός μολυσμένων τερματικών,η αναβάθμιση ή η επιδιόρθωση τρωτών σημείων ενός ευάλωτου τερματικού,το μπλοκάρισμα της κίνησης του σκουληκιού μέσω τοίχων προστασίας ή δρομολογητών,η αποσύνδεση των τερματικών από το δίκτυο.

  β)Ο ρυθμός μόλυνσης β(t) δεν είναι σταθερός.Όσο περνάει ο χρόνος από την απελευθέρωση του σκουληκιού,τόσο ο ρυθμός μόλυνσης μειώνεται.Αυτό δικαιολογείται από το γεγονός ότι ,κατά τη διάδοση ενός σκουληκιού σε ένα δίκτυο μεγάλης κλίμακας,η κίνηση αυξάνεται κατά πολύ με αποτέλεσμα την υπερφόρτωσή του.

- $$\frac{dS(t)}{dt} = -\beta(t)S(t)I(t) - \frac{dQ(t)}{dt}$$
  $$\frac{dR(t)}{dt} = \gamma I(t)$$
  $$\frac{dQ(t)}{dt} = \mu S(t)J(t)$$
  $$\beta(t) = \beta_0[1 - \frac{I(t)}{N}]^n$$
  $$N = S(t) + I(t) + R(t) + Q(t)$$
  $$I(0) = I_0 << N \ , \ S(0) = N - I_0 \ , \ R(0) = Q(0) = 0$$

- Συνεχούς χρόνου,ρεαλιστικό,πολύ ακριβές σε δίκτυα μεγάλης κλίμακας.

<u>Μοντέλο Compartment-based:</u>

- Serazzi και Zanero[28]
- Το Internet προσεγγίστηκε μακροσκοπικά σαν διασύνδεση ενός αριθμού αυτόνομων συστημάτων (Autonomous systems-AS),τα οποία ουσιαστικά είναι υποδίκτυα που διαχειρίζονται από μία και μόνο αρχή(authority)
- Εξέταση εξάπλωσης: α) στο εσωτερικό ενός αυτόνομου συστήματος (ή στο εσωτερικό μίας περιοχής εντός του αυτόνομου συστήματος,η οποία χαρακτηρίζεται από πυκνές συνδέσεις) όπου το σκουλήκι διαδίδεται ανεμπόδιστο και η διάδοση του ακολουθεί το μοντέλο RCS.

  β) ανάμεσα στα αυτόνομα συστήματα

  $$\frac{da_1}{dt} = [a_1 KP_{IN,1} + \frac{N_2}{N_1} a_2 KP_{OUT,2}](1 - a_1)$$
  $$\frac{da_2}{dt} = [a_2 KP_{IN,2} + \frac{N_1}{N_2} a_1 KP_{OUT,1}](1 - a_2)$$

  Με $P_{IN,1}=N_1/N$ και $P_{OUT,1}=1-P_{IN,1}=N_2/N$ θεωρώντας ότι το σκουλήκι γεννά τυχαία την διεύθυνση IP του στόχου.

4 Μοντέλα εξάπλωσης σκουληκιών σε πραγματικά δίκτυα:

4.1 Μοντέλα εξάπλωσης σκουληκιών μαζικής αποστολής με mailing list σε email δίκτυα:

Το ηλεκτρονικό ταχυδρομείο (email) στις μέρες μας έχει γίνει αναπόσπαστο κομμάτι της καθημερινής μας επικοινωνίας.Αυτός είναι και ο κύριος λόγος για τον οποίο κάνουν την εμφάνισή τους όλο και περισσότερα κακόβουλα προγράμματα τα οποία το χρησιμοποιούν σαν μέσο για την εξάπλωση τους και έχουν σαν σκοπό την μόλυνση των υπολογιστών ανυποψίαστων χρηστών.

Ένα είδος τέτοιων προγραμμάτων είναι τα σκουλήκια μαζικής αποστολής (mass-mailing) τα οποία κάνουν χρήση λίστας αποστολής (mailing-list) με email διευθύνσεις για να επιταχύνουν τη διαδικασία εξάπλωσής τους.

Στην παρούσα φάση της εργασίας θα αναλυθεί ένας μηχανισμός διάδοσης

τέτοιου είδους σκουληκιών και θα αποδειχθεί ότι η ύπαρξη λίστας αποστολής επιρρεάζει τη διάδοση τους περισσότερο απ'ότι η τοπολογία του δικτύου.Αρχικά όμως θα αναφερθούν κάποια χαρακτηριστικά των παραπάνω σκουληκιών καθώς και των email δικτύων[29].

4.1.1 Σκουλήκια μαζικής αποστολής (mass-mailing worms):

Τα σκουλήκια μαζικής αποστολής συνήθως περιέχονται σε αρχεία τα οποία έχουν επισυναφθεί σε email.Συνεπώς για να αρχίσει η διαδικασία μόλυνσης ενός υπολογιστή από ένα τέτοιου είδους σκουλήκι θα πρέπει ο χρήστης αρχικά να λάβει ένα τέτοιο email και να εκτελέσει το αρχείο μέσα στο οποίο βρίσκεται το κακόβουλο πρόγραμμα.Από τη στιγμή που συμβεί αυτό,ο υπολογιστής έχει μολυνθεί.Στη συνέχεια το σκουλήκι συλλέγει email διευθύνσεις είτε από το βιβλίο διευθύνσεων του χρήστη είτε από ολόκληρο τον σκληρό δίσκο του υπολογιστή και με αυτόν τον τρόπο δημιουργείται η λίστα αποστολής του σκουληκιού.Βασική στρατηγική των σκουληκιών μαζικής αποστολής είναι να στείλουν μολυσματικά emails σε όλες τις διευθύνσεις που συλλέχθηκαν.Πλέον,αυτό το είδος σκουληκιών χρησιμοποιεί βελτιωμένη λίστα διευθύνσεων,προσθέτωντας γνωστά ονόματα λογαριασμών και απομακρύνοντας διευθύνσεις με κυβερνητικά domains ή domains εταιρειών που σχετίζονται με την ασφάλεια συστημάτων.

Ο μολυσματικός υπολογιστής συνήθως ξεκινά κάνοντας μία επίθεση άρνησης παροχής υπηρεσιών (Denial of Service-DoS) και από τη στιγμή που μία τέτοια επίθεση διενεργηθεί από μεγάλο αριθμό μολυσματικών υπολογιστών θα καταλήξει σε επίθεση κατανεμημένης άρνησης παροχής υπηρεσιών(Distributed Denial of Service-DDoS).

4.1.2 Email Δίκτυο:

Από τον μηχανισμό μόλυνσης που περιγράφηκε παραπάνω συνεπάγεται ότι η διάδοση εξαρτάται από την τοπολογία του δικτύου μέσω του οποίου στέλνεται το email και στο οποίο αναφερόμαστε σαν email δίκτυο.

Πρόσφατες έρευνες στα πολύπλοκα δίκτυα έχουν δείξει ότι τα email δίκτυα έχουν συγκεκριμένα χαρακτηριστικά.Οι Ebel et al. απέσπασαν από τα log αρχεία ενός SMTP εξυπηρετητή,τα ζευγάρια "From" και "To" των email διευθύνσεων και ανέλυσαν το δίκτυο.Θεωρώντας λοιπόν,τις email διεθύνσεις σαν κόμβους και τα ζευγάρια "From-To" σαν σύνδεσμους,τότε η κατανομή βαθμού έχει scale-free ιδιότητα[30].

Στο συγκεκριμένο μοντέλο το δίκτυο email θα θεωρηθεί scale-free και για την μοντελοποίηση θα χρησιμοποιηθεί μία τροποποιημένη εκδοχή του κλασσικού επιδημιολογικού μοντέλου.

Πιο συγκεκριμένα,το κλασσικό επιδημιολογικό μοντέλο χρειάζεται κάποιες τροποποιήσεις για να μπορέσει να περιγράψει την διαδικασία εξάπλωσης ενός σκουληκιού μαζικής αποστολής και να δείξει τον βαθμό επιρροής της λίστας αποστολής στη διάδοση του σκουληκιού.

Με άλλα λόγια θα πρέπει να ικανοποιούνται οι παρακάτω προυποθέσεις:

α) Τα email δίκτυα έχουν θεωρηθεί scale-free κατά τη μελέτη διάδοσης σκουληκιών μαζικής αποστολής.Παρ'όλα αυτά μερικές μελέτες (Newman et al.[31]) έδειξαν ότι η κατανομή βαθμού σε τέτοια δίκτυα είναι εκθετική.Συνεπώς για να γίνει ανάλυση της διάδοσης ενός τέτοιου σκουληκιού σε πραγματικό δίκτυο,πρέπει το μοντέλο να μπορεί να προσαρμοστεί στην οποιαδήποτε τοπολογία.

β) Στα κλασσικά επιδημιολογικά μοντέλά ένας κόμβος μπορεί να μεταβεί από την ευάλωτη(S) κατάσταση στην μολυσματική(I) και από την μολυσματική στην απομακρυσμένη(R).Με την παρουσία όμως αντιιικού λογισμικού είναι απαραίτητο να συμπεριληφθεί στο μοντέλο και η μετάβαση από την ευάλωτη κατάσταση στην απομακρυσμένη.

γ) Οι email διευθύνσεις μπορούν να θεωρηθούν σαν κόμβοι σε ένα δίκτυο.Παρ'όλα αυτά όταν υπάρχει λίστα αποστολής,η συμπεριφορά της είναι τελείως διαφορετική.Μία λίστα αποστολής λειτουργεί σαν ένα είδος "ενισχυτή" για το email μήνυμα,είτε αυτό περιέχει ένα σκουλήκι είτε όχι.Διαισθητικά τέτοιου είδους "ενίσχυση" επιταχύνει την διάδοση του σκουληκιού και συνεπώς είναι σημαντικό να ληφθεί σοβαρά υπόψη κατά την μοντελοποίηση.

4.1.3 Παραλλαγή SI/SIR μοντέλου με χρήση πινάκων γειτνίασης:

Για την ικανοποίηση των παραπάνω προυποθέσεων χρησιμοποιείται μία παραλλαγή του κλασσικού SI/SIR μοντέλου στην οποία γίνεται χρήση πινάκων γειτνίασης. Για να γίνει αυτό θα χρησιμοποιηθεί η παρακάτω σημειογραφία:

α) Γενικά:

$A=\{\alpha_{xy}\}$=MxN πίνακας A και τα στοιχεία του.

E μοναδιαίος πίνακας

AB πολ/σμος πινάκων και A*B πολ/σμος στοιχείων πινάκων και

$F(A)=\{f(\alpha_{xy})\}$ όπου

$$f(\alpha_{xy})=\begin{cases}1 \text{ αν } \alpha_{xy} >= 0.5\\ 0 \text{ διαφορετικά}\end{cases}$$

β)Πίνακες Καταστάσεων:

$$I(t)=\{i_x(t)\} \text{ όπου } i_x(t)=\begin{cases}1 & \text{αν ο κόμβος x είναι μολυσματικός σε χρόνο t}\\ 0 & \text{διαφορετικά}\end{cases}$$

$$R(t)=\{r_x(t)\} \text{ όπου } r_x(t)=\begin{cases}1 & \text{αν ο κόμβος x είναι απομακρυσμένος σε χρόνο t}\\ 0 & \text{διαφορετικά}\end{cases}$$

Άν $i_x(t)=0$ και $r_x(t)=0$ τότε ο κόμβος x βρίσκεται στην κατάσταση S δηλαδή είναι ευάλωτος.

γ)Δίκτυο:

$$T=\{t_{xy}\} \text{ όπου } t_{xy}=\begin{cases}1 & \text{αν από τον κόμβο x στον y υπάρχει σύνδεσμος}\\ 0 & \text{διαφορετικά}\end{cases}$$

Άν ο πίνακας Τ είναι συμμετρικός δηλαδή $t_{xy}=t_{yx}$ τότε το δίκτυο που ορίζεται από τον Τ είναι ένας μη κατευθυνόμενος γράφος.

Μετάβαση κατάστασης του R(t):

$R(t+1)=F(R(t)+I(t)*\Lambda)$ και $r_x(t+1)=f(r_x(t)+i_x(t)*\lambda_x(t))$

$$\lambda_x(t)=\begin{cases}1 & ,\text{αν η τυχαία μεταβλητή } \tau<=\gamma \text{ (ρυθμός απομάκρυνσης)}\\ 0 & ,\text{διαφορετικά}\end{cases}$$

όπου $\Lambda=\{\lambda_x(t)\}$

Μετάβαση κατάστασης του I(t):

$I(t+1)=F(I(t)(E+D(t)*T))*(1-R(t))$ και $i_x(t+1)=f(\Sigma i_y(t)(e_{yx}+d_{yx}(t)*t_{yx}))(1-r_x(t))$

$$d_{xy}(t)=\begin{cases}1 & ,\text{αν η τυχαία μεταβλητή } \tau<=\beta \text{ (ρυθμός μόλυνσης)}\\ 0 & ,\text{διαφορετικά}\end{cases}$$

όπου $D=\{d_{xy}(t)\}$

Παρακάτω παρουσιάζεται το κλασσικό επιδημιολογικο μοντέλο με βάση την παραπάνω σημειογραφία

Έστω S(t) ο αριθμός των κόμβων που βρίσκονται στην κατάσταση S και R(t),I(t) αντίστοιχα για τις καταστάσεις R και I.Οι μεταβάσεις ανάμεσα στις καταστάσεις μπορούν να εκφραστούν ως εξής(1):

$$S(t+1) = S(t) - S(t)\,p(S \to I)$$
$$I(t+1) = I(t) + S(t)\,p(S \to I) - I(t)\,p(I \to R)$$
$$R(t+1) = R(t) + I(t)\,p(I \to R)$$

Η μετάβαση S->I σημαίνει ότι ($i_x(t)=0$ και $\lambda_x(t)=0$)->($i_x(t+1)=1$ και $\lambda_x(t+1)=0$) το οποίο συνεπάγεται ($i_k(t)*d_{kx}(t)*t_{kx}=1$ για $1<=k<=N$)

Άρα για το p(S->I) ισχύει:

$$p(S \to I) = p(\text{οποιοδηποτε } i_y(t) * d_{yx}(t) * t_{yx} = 1) =$$
$$= 1 - (1-\beta)^{\sum_y i_y(t) t_{yx}} = 1 - (1-\beta)^{m_x(t)}$$
$$\text{όπου } m_x(t) = \sum_y i_y(t) t_{yx}$$

με $m_x(t)$ τον αριθμό των μολυσματικών κόμβων που συνδέονται με τον x.

Στην περίπτωση πλήρους δικτύου δηλαδή $t_{kx}=1$ για κάθε $k \in [1, N]$ ισχύει $m_x(t)=\Sigma i_y(t)=I(t)$ και αν $\beta<<1$ τότε $p(S->I)=\beta I(t)$ και $p(I->R)=\gamma I(t)$ και συνεπώς οι εξισώσεις (1) μπορούν να γραφτούν ως εξής:

$$S(t+1) = S(t) - \beta S(t) I(t)$$
$$I(t+1) = I(t) + \beta S(t) I(t) - \gamma I(t)$$
$$R(t+1) = R(t) + \gamma I(t)$$

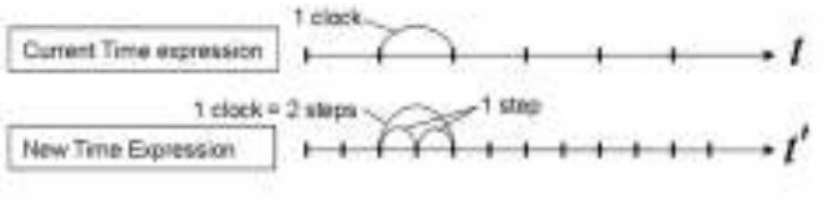
και συγκεκριμένα αν η μετάβαση κατάστασης I(t+1) γραφτεί σαν διαφορική εξίσωση τότε προκύπτει η σχέση $dI/dt=\beta SI-\gamma I$ του κλασσικού μοντέλου SIR.Το παραπάνω μοντέλο ικανοποιεί την πρώτη προυπόθεση που τέθηκε προηγουμένως.

Για να ικανοποιηθεί η δεύτερη θα χρειαστεί απλά να αλλαχθεί η εξίσωση $R(t+1)=F(R(t)+I(t)*\Lambda)$ και να γίνει $R(t+1)=F(R(t)+\Lambda)$.

Τώρα για να ικανοποιηθεί η τρίτη προυπόθεση θα χρειαστεί επέκταση της σημειογραφίας που χρησιμοποιήθηκε παραπάνω.Μία λίστα αποστολής (από δω και στο εξής θα αναφέρεται ως ML-mailing list) συμπεριφέρεται σαν ειδικός κόμβος.Όταν ένας ML κόμβος λαμβάνει ένα email από έναν άλλο κόμβο ,το προωθεί κατ'ευθείαν σε όλα τα μέλη της ML.Από την άλλη όταν ένας κανονικός κόμβος (χρήστης email) λαμβάνει ένα email (στη συγκεκριμένη περίπτωση ένα email που περιέχει ένα σκουλήκι μαζικής αποστολής) ,το email δεν προωθείται σε άλλους κόμβους μέχρι ο χρήστης που το έλαβε να μολυνθεί,δηλαδή μέχρι να ενεργοποιησεί την μολυσματική επισύναψη του μηνύματος.

Ένας ML κόμβος έχει δύο στάδια δράσης ανά χρονική στιγμή (μόλυνση και

διάδοση) ενώ ένας κανονικός κόμβος μόνο ένα στάδιο δράσης ανά χρονική στιγμή. Για να γίνουν εμφανή τα ειδικά χαρακτηριστικά ενός ML κόμβου χρειάζεται μία τροποποίηση της χρονικής κλίμακας.Όταν η τιμή του χρόνου είναι άρτιος αριθμός,διενεργούνται το στάδιο δράσης του κανονικού κόμβου και το πρώτο στάδιο δράσης του ML κόμβου (μόλυνση).Το δεύτερο στάδιο δράσης του ML κόμβου (διάδοση) διενεργείται στις περιττές τιμές του χρόνου(Σχ.1).



Σχ.1 Τροποποιημένη χρονική κλίμακα μοντέλου

Στη συνέχεια γίνεται επέκταση της σημειογραφίας ώστε να συμπεριληφθούν και οι ML κόμβοι.

Οι πίνακες καταστάσεων I(t) και R(t) καθώς και η έκφραση του δικτύου T επεκτείνονται σε I'(t),R'(t) και T' αντίστοιχα,χρησιμοποιώντας τους πίνακες καταστάσεων $I_{\{ML\}}(t)$ και $R_{\{ML\}}(t)$ και την τοπολογία δικτύου $T_{\{U->ML\}}$, $T_{\{ML->U\}}$ και $T_{\{ML->ML\}}$ που εκφράζουν τις τοπολογίες μεταξύ κανονικών κόμβων (κόμβων χρήστη) και ML κόμβων.

$$I'(t) = \{I(t), I_{\{ML\}}(t)\} \qquad R'(t) = \{R(t), R_{\{ML\}}(t)\}$$

$$T' = \begin{pmatrix} T & T_{\{U \circledR ML\}} \\ T_{\{ML \circledR U\}} & T_{\{ML \circledR ML\}} \end{pmatrix} \qquad \hat{T}_{\{U\circledR\}} = \{T, T_{\{U\circledR ML\}}\} \quad \hat{T}_{\{ML\circledR\}} = \{T_{\{ML\circledR U\}}, T_{\{ML\circledR ML\}}\}$$

Στη συνέχεια γίνεται επέκταση των πινάκων καταστάσεων R'(t) και I'(t).

$$R'(2t+1) = F(R'(2t) + \{\Lambda, 0\}) \text{ και } R'(2t+2) = R'(2t+1)$$

$$I'(2t+1) = F(I'(2t)(\{E,0\} + D'(2t) * \hat{T}_{\{U\to\}})) * (\{1 - R'(2t), 0\}) \text{ και}$$
$$I'(2t+2) = F(\{I'(2t+1), 0\} + I_{\{ML\}}(2t+1) * \hat{T}_{\{ML\to\}}) * (1 - R'(2t+2))$$

Τέλος παρέχονται οι διαφορικές εξισώσεις για το τροποποιημένο SIR μοντέλο και για το μοντέλο λίστας αποστολής.

Τροποποιημένο SIR μοντέλο για σκουλήκια μαζικής αποστολής

$$\frac{dS}{dt} = -(1-\gamma)\beta S m_x(t) - \gamma S$$
$$\frac{dI}{dt} = (1-\gamma)\beta S m_x(t) - \gamma I$$
$$\frac{dR}{dt} = \gamma(S+I)$$

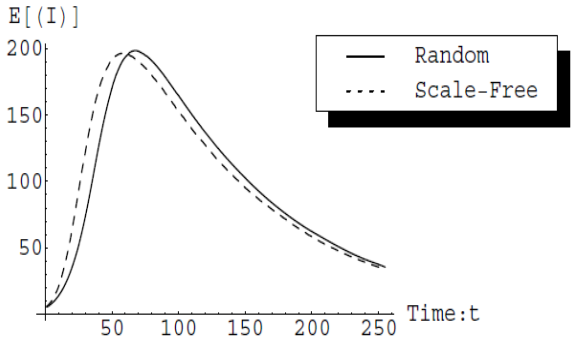
Μοντέλο λίστας αποστολής

$$\frac{dS}{dt} = -(1-\gamma)\beta S\Omega - \gamma S$$
$$\frac{dI}{dt} = (1-\gamma)\beta S\Omega - \gamma I$$
$$\frac{dR}{dt} = \gamma(S+I)$$
$$\text{όπου} \quad \Omega = m_x(2t) + \sum_{y=N+1}^{N+M} m_y(2t)t_{yx}$$

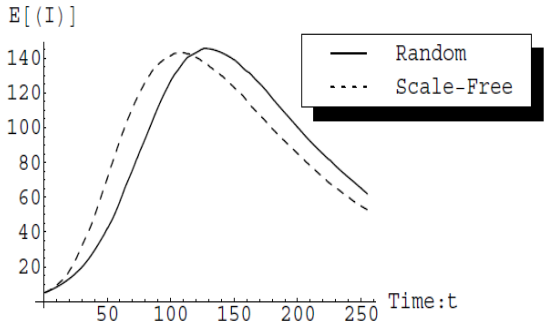
4.1.4 Προσομοιώσεις-Συμπεράσματα:

Τέλος, με χρήση προσομοιώσεων θα γίνει εμφανές ότι η λίστα αποστολής έχει μεγαλύτερη επιρροή στη διάδοση του σκουληκιού από την τοπολογία του δικτύου. Στα πλαίσια όλων των προσομοιώσεων ο αριθμός των κόμβων έχει θεωρηθεί N=300 και ο αριθμός των αρχικά μολυσματικών κόμβων I(0)=5.Επιπλέον,έχει

χρησιμοποιηθεί ο αλγόριθμος-γεννήτρια scale-free τοπολογίας από την μελέτη των Barabasi και Albert [5].Έχει θεωρηθεί ότι $T_{\{ML->ML\}}=0$ που σημαίνει ότι τα μέλη μίας λίστας αποστολής δεν περιέχονται σε άλλες λίστες αποστολής.



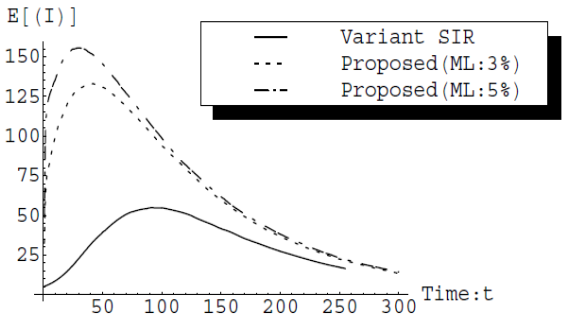
Σχ.2 SIR μοντέλο με β=0.005,γ=0.01 E[K]=12



Σχ.3 SIR μοντέλο με β=0.01,γ=0.01 E[K]=12

Αρχικά εξετάζονται οι διαφορές ανάμεσα στην διάδοση σε scale-free δίκτυα και σε τυχαίους γράφους. Χρησιμοποιώντας το κλασσικό μοντέλο SIR,δείχνουμε τις μεταβολές στην τιμή I(t) των μολυσματικών κόμβων για κάθε μία από τις δύο περιπτώσεις, θεωρώντας ρυθμό μόλυνσης β=0.005,ρυθμό απομάκρυνσης γ=0.01 και μέσο βαθμό όλου του δικτύου E[K]=12 για το Σχ.2 και β=0.01,γ=0.01 και E[K]=12 για το Σχ.3.

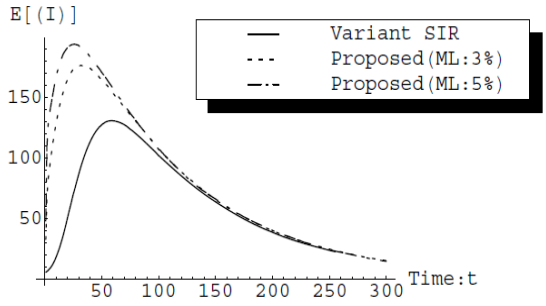
Παρατηρώντας τις γραφικές παραστάσεις των Σχ.2 και Σχ.3 είναι εμφανές ότι τα σκουλήκια εξαπλώνονται γρηγορότερα μέσω scale-free δικτύου από ότι μέσω τυχαίου γράφου και η μεγιστοποίηση του I(t) συμβαίνει νωρίτερα στα πρώτα.Στη συνέχεια θα γίνει σύγκριση μεταξύ του μοντέλου ML και του τροποποιημένου SIR μοντέλου με σκοπό να γίνει εμφανής η επιρροή της λίστας αποστολής στη διάδοση του σκουληκιού.



Σχ.4 Επίδραση ML, β=0.005,γ=0.01

Για τις ανάγκες της προσομοίωσης ετοιμάστηκαν δύο δίκτυα με ML κόμβους.Το πρώτο περιέχει 3% ML κόμβους και το δεύτερο 5%.Και στις δύο περιπτώσεις E[K]=12 και η τοπολογία του δικτύου είναι scale-free.Το Σχ.4 δείχνει το αποτέλεσμα για β=0.005 και γ=0.01 (Περιπτωση 1) και το Σχ.5 δείχνει το

αποτέλεσμα για β=0.01 και γ=0.01(Περιπτωση 2).Παρατηρείται αρκετά γρήγορη διάδοση και υψηλή κορυφή και στις δύο περιπτώσεις.



Σχ.5 Επίδραση ML, β=0.01,γ=0.01

Αξίζει να σημειωθεί ότι η διαφοροποίηση στις καμπύλες των Σχ.4 και Σχ.5 είναι αρκετά μεγάλη ειδικά ανάμεσα στην περίπτωση που υπάρχει λίστα αποστολής και στην περίπτωση που δεν υπάρχει.Κάτι τέτοιο δεν συμβαίνει στα Σχ.2 και Σχ.3 που αλλάζει η τοπολογία του δικτύου.

Συνεπώς είναι εμφανές ότι η ύπαρξη λίστας αποστολής επιρρεάζει σε μεγάλο βαθμό την διάδοση του σκουληκιού σε αντίθεση με την εναλλαγή στην τοπολογία του δικτύου από scale-free σε τυχαίου γράφου.

4.1.5. Σύνοψη:

- Kanaoka, Okamoto (2008 [29])
- Σκουλήκια μαζικής αποστολής:συνήθως περιέχονται σε αρχεία τα οποία έχουν επισυναφθεί σε email.Όταν εκτελεστεί το αρχείο,μολύνεται ο υπολογιστής.Συλλέγονται διευθύνσεις email για την δημιουργία της λίστας αποστολής και γίνεται αποστολή του μολυσματικού email σε όλες τις διευθύνσεις της λίστας.
- Χαρακτηριστικά μοντέλου:

  Α)ανεξάρτητο τοπολογίας δικτύου(εκθετική ή power law κατανομή βαθμού)

  Β)τροποποιημένο SIR μοντέλο-λαμβάνεται υπόψιν η πιθανή παρουσία αντιικού λογισμικού.

  Γ)κόμβοι του δικτύου είναι οι email διευθύνσεις.Λαμβάνεται υπόψιν η επιρροή της λίστας αποστολής στην εξάπλωση του σκουληκιού,η οποία λειτουργεί σαν ένα είδος “ενισχυτή” για το email μήνυμα.

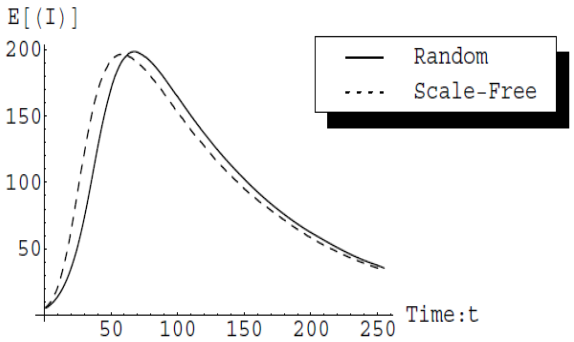
- Τροποποιημένο SIR μοντέλο για σκουλήκια μαζικής αποστολής

$$\frac{dS}{dt} = -(1-\gamma)\beta S m_x(t) - \gamma S$$
$$\frac{dI}{dt} = (1-\gamma)\beta S m_x(t) - \gamma I$$
$$\frac{dR}{dt} = \gamma(S+I)$$

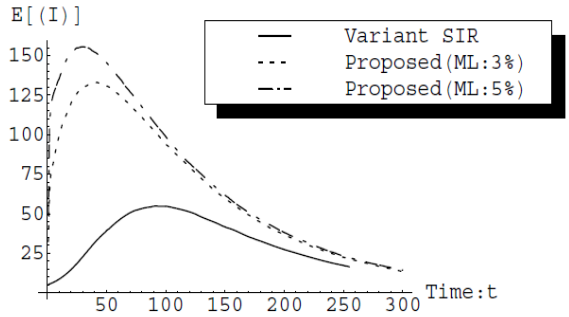
- Προτεινόμενο μοντέλο λίστας αποστολής.

$$\frac{dS}{dt} = -(1-\gamma)\beta S\Omega - \gamma S$$
$$\frac{dI}{dt} = (1-\gamma)\beta S\Omega - \gamma I$$
$$\frac{dR}{dt} = \gamma(S+I)$$
$$\text{όπου} \quad \Omega = m_x(2t) + \sum_{y=N+1}^{N+M} m_y(2t)t_{yx}$$

- Γρηγορότερη εξάπλωση μέσω scale-free δικτύου από ότι μέσω τυχαίου γράφου και η μεγιστοποίηση του I(t) συμβαίνει νωρίτερα στο πρώτο.Χρήση κλασσικού SIR μοντέλου.



- Όσο μεγαλύτερο είναι το ποσοστό ML κόμβων στο δίκτυο,η ταχύτητα εξάπλωσης του σκουληκιού και το μέγιστο πλήθος μολυσματικών κόμβων αυξάνονται.



- Η μεγαλύτερη διαφοροποίηση στην εξάπλωση, εντοπίζεται κατά τη σύγκριση του τροποποιημένου SIR με το προτεινόμενο μοντέλο με ύπαρξη ML. είναι εμφανές ότι η ύπαρξη λίστας αποστολής επιρρεάζει σε μεγάλο βαθμό την διάδοση του σκουληκιού σε αντίθεση με την εναλλαγή στην τοπολογία του δικτύου από scale-free σε τυχαίου γράφου.

4.2 Μοντέλο εξάπλωσης ενεργών σκουληκιών σε p2p δίκτυα:

Όπως εχουμε ήδη αναφέρει σε προηγούμενο κεφάλαιο τα ενεργά σκουλήκια στις μέρες μας, αποτελούν τεράστια απειλή για την ασφάλεια του Internet.Εξαπλώνονται μόνα τους, αντιγράφονται μόνα τους και συνεπώς έχουν τη δυνατότητα να διαδίδονται μέσω του εκάστοτε δικτύου, από μολυσματικούς κόμβους σε ευάλωτους,χωρίς καμία παρέμβαση από τον χρήστη.Επιπλέον τα p2p συστήματα έχουν γίνει ιδιαίτερα δημοφιλή ανάμεσα στους χρήστες του Internet.Έχοντας χαρακτηριστικά όπως εξυπηρέτηση μεγάλου αριθμού χρηστών και υψηλή συνδετικότητα,τα p2p συστήματα προτιμώνται για την απελευθέρωση ενεργών σκουληκιών(από τον δημιουργό κακόβουλου λογισμικού)μίας και μέσω αυτών επιτυγχάνεται γρηγορότερη διάδοση.

Πιο αναλυτικά,οι παρακάτω διαπιστώσεις κάνουν εμφανείς τους λόγους για τους οποίους η διάδοση ενός σκουληκιού μέσω p2p συστημάτων διευκολύνει τις επιθέσεις του σε ευάλωτους κόμβους.

α) Αρχικά τα p2p συστήματα,όπως αναφέραμε και προηγουμένως,περιλαμβάνουν ένα μεγάλο αριθμό εγγεγραμμένων τερματικών που σημαίνει ότι κατά την διαδικασία

μόλυνσης ενός τέτοιου συστήματος,η διάδοση του σκουληκιού θα επιταχυνθεί αφού τα τερματικά αυτά είναι πραγματικά και ενεργά.

β) Επιπλέον από τη στιγμή που τα τερματικά σε αυτά τα συστήματα διατηρούν έναν μεγάλο αριθμό γειτόνων για λόγους δρομολόγησης,είναι φανερό ότι το μολυσμένο τερματικό θα μπορέσει με ευκολία να διαδώσει το σκουλήκι στους γειτονές του.

γ) Μερικά τερματικά p2p συστημάτων είναι πιθανόν να μην ικανοποιούν βασικές αρχές ασφάλειας όπως για παραδειγμα ένα οικιακό δίκτυο.

δ) Τέλος, πολλοί χρήστες εγκαθιστούν και χρησιμοποιούν p2p προγράμματα,τα οποία λόγω κάποιων τρωτών σημείων τους διευκολύνουν την διαδικασία μόλυνσης και συνεπώς την εξάπλωση του σκουληκιού.

Παρακάτω θα καθοριστούν δύο μοντέλα επίθεσης p2p-based σκουληκιών.Ένα μοντέλο offline p2p based επίθεσης με χρήση hitlist και ένα online p2p based επίθεσης.

4.2.1 Μοντέλα επιθέσεων p2p based σκουληκιών:

Ένα ενεργό σκουλήκι είναι ένα κακόβουλο πρόγραμμα που διαδίδεται από τερματικό σε τερματικό,μέσω δικτύου, εκμεταλλευόμενο τρωτά σημεία στην ασφάλεια αυτών.

Γενικά υπάρχουν δύο στάδια στην επίθεση ενός ενεργού σκουληκιού.Η σάρωση του δικτύου για εύρεση πιθανών στόχων(τερματικών) και η μόλυνση αυτών αφού διαπιστωθεί ότι είναι ευάλωτοι.Στα παραπάνω δύο στάδια καθοριστικό ρόλο στην ταχύτητα εξάπλωσης του σκουληκιού παίζει η ταχύτητα με την οποία ανιχνεύει άλλα τερματικά στο δίκτυο,η πιθανότητα να ανιχνεύσει ένα πραγματικό τερματικό και το αν το ανιχνευμένο τερματικό είναι ευάλωτο ή όχι.

Πιο συγκεκριμένα,η ταχύτητα με την οποία το σκουλήκι ανιχνεύει άλλα τερματικά μοντελοποιείται από τον ρυθμό σάρωσης S,που δείχνει τον αριθμό των τερματικών ανά μονάδα χρόνου που ένα μολυσματικό τερματικό μπορεί να ανιχνεύσει.Η πιθανότητα ανίχνευσης ενός πραγματικού τερματικού σχετίζεται με το γεγονός ότι μόνο το 24% των διευθύνσεων του Internet χρησιμοποιείται από ενεργά τερματικά.Παρόλα αυτά κατά τη διάδοση σε p2p συστήματα,η σάρωση είναι πιο ακριβής αφού όπως προαναφέρθηκε,τα p2p συστήματα έχουν μεγάλο αριθμό

πραγματικών και ενεργών τερματικών με μεγάλη συνδετικότητα μεταξύ τους.Τέλος όσων αφορά στο αν το ανιχνευμένο τερματικό είναι ευάλωτο ή όχι,η πιθανότητα ανίχνευσης ενός ευάλωτου τερματικού στα συγκεκριμένα συστήματα είναι αρκετά υψηλή καθώς τα περισσότερα τερματικά που εισάγονται σε p2p συστήματα είναι συνήθως μη έμπιστα και μη έγκυρα.

Στη συνέχεια παρουσιάζονται τα δύο μοντέλα p2p based επίθεσης όπου το σκουλήκι διαδίδεται μέσω p2p συστημάτων για την επίτευξη γρηγορότερης εξάπλωσης.

4.2.1.1 *Offline p2p based hit-list scan(OPHLS):*

Στο συγκεκριμένο μοντέλο,ο επιτιθέμενος συλλέγει πληροφορίες για IP διευθύνσεις από το p2p σύστημα ενώ είναι εκτός σύνδεσης(offline).Η συλλογή αυτή των IP διευθύνσεων αποτελεί την hitlist του σκουληκιού.

Το συγκεκριμένο μοντέλο χωρίζεται σε δύο φάσεις.Κατά την πρώτη φάση(φάση επίθεσης)όλα τα τερματικά που μόλις μολύνθηκαν εξαπολύουν επίθεσεις στα τερματικά των οποίων οι IP διευθύνσεις βρίσκονται στην hitlist.Οι επιθέσεις δεν σταματάνε μέχρι να έχουν σαρωθεί όλα τα τερματικά της hitlist.Στη δεύτερη φάση,όλα τα μολυσματικά τερματικά συνεχίζουν τις επιθέσεις στο Internet χρησιμοποιώντας μηχανισμό τυχαίας σάρωσης.

4.2.1.2 *Online p2p based scan(OPS):*

Στο συγκεκριμένο μοντέλο,τα σκουλήκια κατά τη διάδοση τους χρησιμοποιούν την υψηλή συνδετικότητα των p2p συστημάτων.Μόλις ένα μολυσματικό τερματικό εισέλθει στο p2p σύστημα,εξαπολύει επίθεση με υψηλή προτεραιότητα στους p2p γείτονές του.Για παράδειγμα θεωρούμε ένα μολυσματικό τερματικό έστω A_1,στο p2p σύστημα,με ρυθμό σάρωσης σκουληκιού S=5.Έστω ότι το A_1 έχει τρεις p2p γείτονες B_1,B_2,B_3.Στο μοντέλο OPS το A_1 θα επιτεθεί στα B_1,B_2 και B_3 χρησιμοποιώντας το 60% των δυνατοτήτων σάρωσης που έχει και με το υπόλοιπο 40% θα επιτεθεί στο Internet χρησιμοποιώντας τυχαία σάρωση.Αν υποτεθεί τώρα ότι τα B_2 και B_3 είναι ευάλωτα τερματικά και μολυνθούν,θα συνεχίσουν τις επιθέσεις με τον ίδιο τρόπο δίνοντας προτεραιότητα στους p2p γείτονες τους.Στο σημείο αυτό το A_1 τερματικό δεν έχει άλλο p2p τερματικό να σαρώσει οπότε επιτίθεται με το 100% των δυνατοτήτων του στο Internet χρησιμοποιώντας

μηχανισμό τυχαίας σάρωσης.

4.2.2 Παράμετροι μοντέλου εξάπλωσης:

Οι παράμετροι που χρησιμοποιούνται για τη μοντελοποίηση της εξάπλωσης των σκουληκιών σε p2p συστήματα μπορούν να διαχωριστούν σε δύο κατηγορίες.Σε αυτές που σχετίζονται με την επίθεση και σε αυτές που σχετίζονται με τα p2p συστήματα.Οι παράμετροι που σχετίζονται με την επίθεση είναι ο ρυθμός σάρωσης S που αναφέρθηκε προηγουμένως,και ο αριθμός των αρχικά μολυσμένων τερματικών M(0).Διαισθητικά υψηλότερες τιμές του S και του M(0) σημαίνει και δυνατότερη επίθεση από πλευράς του σκουληκιού.Οι παράμετροι που σχετίζονται με τα p2p συστήματα είναι το μέγεθος του συστήματος,η δομημένη/μη δομημένη τοπολογία του και η ευαισθησία των p2p τερματικών.Πιο αναλυτικα:Σαν μέγεθος του συστήματος ορίζεται ο αριθμός των χρηστών ενός p2p συστήματος.Στην πραγματικότητα υπάρχουν πολλά διαφορετικά p2p συστήματα.Παρόλα αυτά για τους σκοπούς της εργασίας θα θεωρήσουμε ένα Super-p2p σύστημα που θεωρητικά περιλαμβάνει όλα τα p2p τερματικά του Internet.Τα υπόλοιπα τερματικά του Internet θεωρούνται μέρος του Non-p2p συστήματος.Το μέγεθος του Super-p2p συστήματος παριστάνεται με m.Όσον αφορά την δομημένη/μη δομημένη τοπολογία του p2p συστήματος,αυτές οι δύο κατηγορίες διαχωρίζουν τα p2p συστήματα με βάση την τοπολογία τους.Στα δομημένα p2p συστήματα όπως το CAN και Chord όλοι οι p2p κόμβοι έχουν τον ίδιο αριθμό γειτόνων (δηλ. βαθμό τοπολογίας) για πιο αποτελεσματική δρομολόγηση. Αντίθετα,στα μη δομημένα p2p συστήματα όπως το Gnutella και το Freenet ο αριθμός των γειτόνων κάθε κόμβου είναι διαφορετικός.Ο βαθμός τοπολογίας έχει σημαντική επιρροή στην επίδοση του μοντέλου επίθεσης σκουληκιών OPS.Στα δομημένα συστήματα,ο βαθμός τοπολογίας θα παριστάνεται με θ και προφανώς θα είναι σταθερός.Στα μη δομημένα,ο βαθμός τοπολογίας ακολουθεί power-law.Συνεπώς η κατανομή βαθμού P(k),που εκφράζει την πιθανότητα ένας τυχαία επιλεγμένος κόμβος να έχει ακριβώς k συνδεσμους,δίνεται από τον τύπο(1):

$$P(k) = C_1 \frac{\omega}{k^{\sigma}}$$

όπου ω είναι η μέση τιμή του βαθμού τοπολογίας,C_1 είναι μία σταθερά για δοσμένο ω και σ ϵ [1,8] ο βαθμός power-law.

Τέλος,τα τερματικά στα p2p συστήματα πρόκειται για πραγματικούς υπολογιστές που πιθανόν να βρίσκονται σε λιγότερο προστατευμένα περιβάλλοντα όπως σχολεία,σπίτια ή δημόσιους χώρους.Επίσης συνήθως εγκαθιστούν συγκεκριμένες p2p εφαρμογές και οποιοδήποτε τρωτό σημείο των εφαρμογών αυτών μπορεί να χρησιμοποιηθεί από το σκουλήκι ώστε να τα μολύνει.Στην συγκεκριμένη εργασία,για την μοντελοποίηση της ευπάθειας θα χρησιμοποιηθεί η έννοια της πιθανότητας.Συγκεκριμένα με P_3 εκφράζεται η πιθανότητα ένα τερματικό στο Super-p2p σύστημα να είναι ευάλωτο σε μόλυνση από ένα σκουλήκι και με P_2 η αντίστοιχη πιθανότητα στο Non-p2p σύστημα.

4.2.3 Ανάλυση εξάπλωσης p2p based σκουληκιών:

Αρχικά υπάρχουν $M^s(0)$ μολυσματικά τερματικά στο Super-p2p σύστημα και $M^n(0)$ μολυσματικά τερματικά στο Non-p2p.Για την περιγραφή της εξάπλωσης του σκουληκιού χρησιμοποιείται το επιδημιολογικό μοντέλο SIR όπου κάθε τερματικό μπορεί να βρίσκεται σε μία εκ των τριών καταστάσεων susceptible(ευάλωτο), infectious (μολυσματικό) ή removed (για τον σκοπό του κεφαλαίου immune-ανοσοποιημένο). Στο συγκεκριμένο μοντέλο ένα ανοσοποιημένο τερματικό είναι αυτό που δεν μπορεί να μολυνθεί από το σκουλήκι.

Παρακάτω παρουσιάζεται η ανάλυση της εξάπλωσης ενός σκουληκιού σε p2p συστήματα.Το βασικό μέγεθος στην ανάλυση αυτή είναι ο αριθμός των τερματικών που μόλις μολύνθηκαν σε βήμα χρόνου i και παριστάνεται με E(i).Αντιστοιχα με $E^s(i)$ και $E^n(i)$ παριστάνονται οι αριθμοί τερματικών που μόλις μολύνθηκαν στο super-p2p και non-p2p σύστημα σε χρόνο i.

4.2.3.1 *Ανάλυση μοντέλου επίθεσης OPHLS:*

Στο μοντέλο επίθεσης OPHLS,ο επιτιθέμενος αρχικά συλλέγει τις IP διευθύνσεις των τερματικών που βρίσκονται στο super-p2p σύστημα και δημιουργεί με αυτές μία hitlist με τα πιθανά θύματα.Όπως είχε αναφερθεί προηγουμένως,η επίθεση OPHLS ενός σκουληκιού χωρίζεται σε δύο στάδια.Στο πρώτο,το σκουλήκι σαρώνει όλα τα τερματικά που περιέχονται στην hitlist και συνεπώς στο super-p2p

σύστημα και αφου σαρωθούν όλα τα τερματικά της hitlist,στο δεύτερο στάδιο όλα τα μολυσματικά τερματικά σαρώνουν το non-p2p συστήμα χρησιμοποιώντας μηχανισμό τυχαίας σάρωσης.

Το παρακάτω θεώρημα[32]μας δίνει τον αριθμό των τερματικών που μόλις μολύνθηκαν σαν αποτέλεσμα μίας OPHLS επίθεσης.

Με u παριστάνεται η πιθανότητα ένα p2p τερματικό να είναι συνδεδεμένο και με T το πλήθος IP διευθύνσεων στο Internet.

Θεώρημα 1:

Σε μοντέλο επίθεσης OPHLS,στο Super-p2p σύστημα τη χρονική στιγμή i,το πλήθος μολυσματικών και ευάλωτων τερματικών είναι $M^s(i)$ και $N^s(i)$ αντίστοιχα.Την επόμενη χρονική στιγμή (i+1) θα υπάρχουν (2):

$$E^s(i+1)=\begin{cases} N^s(i)[1-(1-\frac{1}{m*u})^{S*M^S(i)}] & ,M^s(i)\le m*u*P_3 \\ 0 & ,M^s(i)>m*u*P_3 \end{cases}$$

τερματικά που μόλις μολύνθηκαν

όπου $M^s(0)=\mathrm{M}_0, N^s(0)=m*u*P_3$

Στο Non-p2p σύστημα,τη χρονική στιγμή i,με πλήθος μολυσματικών και ευάλωτων τερματικών $M^n(i)$ και $N^n(i)$ αντίστοιχα,την επόμενη χρονική στιγμή θα υπάρχουν (3):

$$E^n(i+1)=\begin{cases} 0 & ,M^s(i)\le m*u*P_3 \\ N^n(i)[1-(1-\frac{1}{T})^{(S*M^n(i)+S*M^s(K)} & ,M^s(i)>m*u*P_3 \end{cases}$$

τερματικά που μόλις μολύνθηκαν

όπου $M^n(0)=0, N^n(K)=T*P_1*P_2-m*u*P_3, M^n(K)=M^s(K-1)$

και $K=\min(i)\ ,\forall i$ που ικανοποιεί την $M^s(i)>m*u*P_3$

Χρησιμοποιώντας το θεώρημα 1 μπορεί να καθοριστεί το πλήθος των μολυσματικών τερματικών στο super-p2p και στο non-p2p σύστημα σε χρόνο i (M(i)) όπως επίσης και ο συνολικός αριθμός ευάλωτων τερματικών (N(i)) στα παραπάνω συστήματα.

Ακολουθούν οι αναδρομικές μέθοδοι υπολογισμού (4) που προκύπτουν από τις εξισώσεις (2) και (3).

$$M^n(i+1) = M^n(i) + E^n(i+1) \ , \ N^n(i+1) = N^n(i) - E^n(i+1) \ ,$$
$$M^s(i+1) = M^s(i) + E^s(i+1) \ , \ N^s(i+1) = N^s(i) - E^s(i+1) \ ,$$
$$M(i) = M^s(i) + M^n(i) \quad , \ N(i) = N^s(i) + N^n(i)$$

Από τις παραπάνω αναδρομικές μεθόδους παρατηρούμε ότι όσο το ενεργό μέγεθος του συστήματος (m*u) αυξάνεται (αύξηση στο m ή στο u ή και στα δύο) , τα $E^s(i)$ και $M^s(i)$ αυξάνονται.Σαν επακόλουθο αυξάνεται και το πλήθος των μολυσματικών κόμβων M(i) που σημαίνει γρηγορότερη εξάπλωση του σκουληκιού.

4.2.3.2 *Ανάλυση μοντέλου επίθεσης OPS:*

Όπως αναφέρθηκε και προηγουμένως,στο μοντέλο επίθεσης OPS η διαδικασία διάδοσης του σκουληκιού,διαφέρει στα δομημένα και στα μη δομημένα p2p συστήματα λόγω των διαφορών στην τοπολογία των δύο συστημάτων.Ο υπολογισμός του πλήθους των τερματικών που μόλις μολύνθηκαν γίνεται όπως στο μοντέλο OPHLS,λαμβάνοντας υπόψιν τον βαθμό τοπολογίας των p2p τερματικών και της πιθανότητας ένα p2p τερματικό να μολυνθεί από τους γείτονές του.

Ακολουθεί το θεώρημα για το μοντέλο επίθεσης OPS[32].

Θεώρημα 2:

Σε μοντέλο επίθεσης OPS,στο Super-p2p σύστημα τη χρονική στιγμή i,το πλήθος μολυσματικών και ευάλωτων τερματικών είναι $M^s(i)$ και $N^s(i)$ αντίστοιχα.Την επόμενη χρονική στιγμή (i+1) θα υπάρχουν (5):

$$E^s(i+1) = N^s(i)\left[1-\left(1-\frac{1}{m*u}\right)^{\left(\left(\sum_{j=1}^{E^s(i)}\min(r_j,S)\right)+\frac{m*u}{T}*S*M^n(i)\right)}\right]$$

τερματικά που μόλις μολύνθηκαν

όπου r_j είναι ο βαθμός τοπολογίας του τερματικού j

και $M^s(0) = \mathrm{M}_0, N^s(0) = m*u*P_3$

Στο Non-p2p σύστημα,τη χρονική στιγμή i,με πλήθος μολυσματικών και ευάλωτων

τερματικών $M^n(i)$ και $N^n(i)$ αντίστοιχα,την επόμενη χρονική στιγμή θα υπάρχουν(6):

$$E^n(i+1) = N^n(i)\left[1-\left(1-\frac{1}{T}\right)^{\left(S*(M^n(i)+M^s(i))-\sum_{j=1}^{E^s(i)}\min(r_j,S)-\frac{m*u}{T}*S*M^n(i)\right)}\right]$$

τερματικά που μόλις μολύνθηκαν

όπου r_j είναι ο βαθμός τοπολογίας του τερματικού j

και $M^n(0)=0, N^n(0)=T*P_1*P_2-m*u*P_3$

Στο παραπάνω θεώρημα οι αναδρομικές μέθοδοι υπολογισμού προκύπτουν με παρόμοιο τρόπο με αυτές του θεωρήματος 1(4).

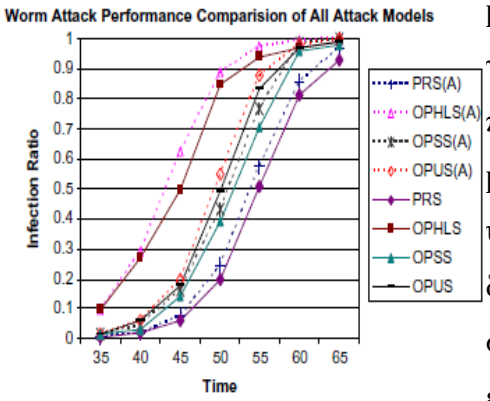
Από τις σχέσεις (5) και (6) παρατηρούμε ότι μία αύξηση του μέσου βαθμού θ (τα δομημένα p2p συστήματα μοντελοποιούνται με βάση την κανονική κατανομή όπου η μέση τιμή των r_j είναι θ) για τα δομημένα p2p συστήματα ή μία αύξηση του μέσου βαθμού ω για τα μη δομημένα p2p συστήματα,προκαλεί την αύξηση του r_j με μεγάλη πιθανότητα.Αυτή η αύξηση του r_j προκαλεί την αύξηση του $E^s(i)$ και την μείωση του $E^n(i)$,με την αύξηση του $E^s(i)$ να είναι μεγαλύτερη από την αντίστοιχη μείωση του $E^n(i)$.Συνεπώς το $E(i)(=E^s(i)+E^n(i))$ αυξάνεται και προκαλεί την αύξηση του συνολικού πλήθους των μολυσματικών τερματικών,άρα η εξάπλωση του σκουληκιού επιταχύνεται.

Παρακάτω θα συγκρίνουμε τα αναλυτικά αποτελέσματα των προηγούμενων υποκεφαλαίων με τα αποτελέσματα προσομοιώσεων ώστε να καταλήξουμε σε συμπεράσματα σχετικά με την επιρροή του μεγέθους και της τοπολογίας ενός p2p συστήματος στην διάδοση ενός σκουληκιού αλλά και σχετικά με την απόδοση κάθε μοντέλου επιθεσης που προτάθηκε παραπάνω.

4.2.4 Προσομοίωση-Συμπεράσματα:

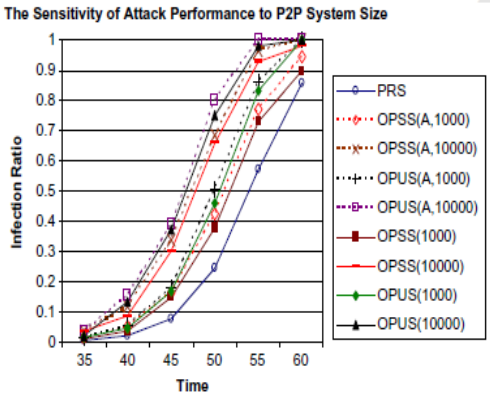
Για την διαδικασία εκτίμησης των μοντέλων επιθέσεων,ορίζεται σαν μέτρο σύγκρισης η αναλογία μόλυνσης ανα μονάδα χρόνου στην οποία αντικατοπτρίζεται η ταχύτητα εξάπλωσης του σκουληκιού.Πιο συγκεκριμένα πρόκειται για την αναλογία του συνολικού αριθμού μολυσματικών τερματικών προς τον αριθμό των ευάλωτων

τερματικών σε βάθος χρόνου.Στην αναλυτική εκτίμηση όπως και στην εκτίμηση προσομοίωσης οι παράμετροι συστήματος που χρησιμοποιούνται είναι $(T,P_1,P_2,P_3,\theta,\sigma,\omega,m,u)=(2^{29},0.25,0.3,0.3,4,3,4,10000,0.7)$ και οι παράμετροι επίθεσης (WA,S,M0)=(*,6,1) όπου WA ϵ {PRS,OPHLS,OPSS,OPUS}[32] ,με * υποδεικνύεται ότι η συγκεκριμένη παράμετρος είναι μεταβλητή στις προσομοιώσεις,με PRS παριστάνεται η τεχνική τυχαίας σάρωσης,με OPSS το μοντέλο OPS για δομημένα p2p συστήματα και με OPUS το μοντέλο OPS για μη δομημένα p2p συστήματα.Στις παρακάτω προσομοιώσεις θεωρείται ότι τα p2p τερματικά έχουν ίδια ευπάθεια με τα υπόλοιπα τερματικά του Internet πράγμα που δεν είναι ρεαλιστικό αφού τα p2p τερματικά στην πραγματικότητα είναι πιο ευπαθή.Συνεπώς,τα αποτελέσματα των προσομοιώσεων θεωρούνται αρκετά αισιόδοξα.



Σχ.6 Επίδοση όλων των μοντέλων επίθεσης

Στο Σχ.6 φαίνονται οι επιδόσεις όλων των μοντέλων που αναφέρθηκαν στο συγκεκριμένο κεφάλαιο συναρτήσει του χρόνου.Ο χρόνος ξεκινάει από την τιμή 35 αφού πριν από εκεί η αναλογία μόλυνσης είναι σχεδόν μηδενική λόγω του πολύ μικρού αριθμού μολυσματικών τερματικών σε σύγκριση με τα ευάλωτα.Το γράμμα Α που βρίσκεται μέσα σε παρένθεση στο υπόμνημα υποδηλώνει ότι τα δεδομένα προέκυψαν από το αντίστοιχο αναλυτικό μοντέλο.Παρατηρώντας τη γραφική παράσταση,συμπεραίνουμε ότι τα p2p based μοντέλα έχουν καλύτερη επίδοση από το PRS μοντέλο όπως επίσης ότι το μοντέλο OPHLS πετυχαίνει γρηγορότερη εξάπλωση σκουληκιού συγκρινόμενο με τα online-based μοντέλα επίθεσης.



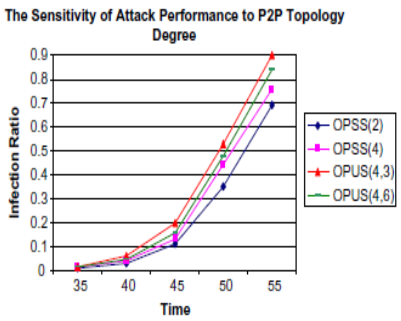
Σχ.7 Επιρροή Μεγέθους P2P συστήματος

Ο λόγος που συμβαίνει το τελευταίο είναι γιατί οι IP διευθύνσεις όλων των p2p τερματικών (hitlist) αποκτώνται πριν ξεκινήσει η επίθεση από το σκουλήκι.

Κατά την προσομοίωση,ο χρόνος που χρειάζεται από τον κακόβουλο χρήστη για την απόκτηση της hitlist δεν λαμβάνεται υπόψη,αφού η συγκεκριμένη δραστηριότητα δεν ανήκει στην διαδικασία εξάπλωσης του σκουληκιου και συνεπως είναι εκτός των πλαισίων της εργασίας.

Στο Σχ.7 φαίνεται η επιρροη του μεγέθους του p2p συστήματος στην αναλογία μόλυνσης και κατ'επέκταση στην εξάπλωση του σκουληκιού.Για τη συγκεκριμένη γραφική παράσταση χρησιμοποιηθηκαν δύο διαφορετικά p2p συστήματα με μεγέθη 1000 και 10000 και παραμέτρους $(T,P_1,P_2,P_3,\theta,\sigma,\omega,m,u)=(2^{29},0.25,0.3,0.3,4,3,4,*,0.7)$. Παρατηρώντας τη γραφική παράσταση γίνεται σαφές ότι όταν το μέγεθος του p2p συστήματος αυξάνεται,η επίδοση όλων των μοντέλων βελτιώνεται.Αυτό συμβαίνει γιατί όταν το μέγεθος του p2p δικτύου αυξάνεται,η πιθανότητα,κατά τη διάρκεια μίας σάρωσης, να βρεθεί ευάλωτο τερματικό αυξάνεται και συνεπώς αυξάνεται και ο αριθμός των μολύνσεων.

Στο Σχ.8 φαίνεται η επιρροή του βαθμού τοπολογίας στην αναλογία μόλυνσης και κατ'επέκταση στην εξάπλωση του σκουληκιού.Οι παράμετροι του συστήματος στην συγκεκριμένη γραφική παράσταση είναι $(T,P_1,P_2,P_3,\theta,\sigma,\omega,m,u)=(2^{29},0.25,0.3,0.3,*,*,10000,0.7)$.Στο υπόμνημα,το OPSS(#) υποδηλώνει το μοντέλο OPSS με βαθμό τοπολογίας # ενώ το OPUS(ω,σ) υποδηλώνει το μοντέλο OPUS με παραμέτρους power-law ω και σ.Παρατηρώντας τη γραφική παράσταση συμπεραίνουμε ότι για το δομημένο p2p σύστημα μία αύξηση στον βαθμό τοπολογίας,επιφέρει γρηγορότερη εξάπλωση σκουληκιού.Αυτό συμβαίνει γιατί όταν ο βαθμός αυξάνεται,αυξάνεται και η συνδετικότητα γεγονός που επιτρέπει στο σκουλήκι την σάρωση περισσότερων τερματικών και συνεπώς γρηγορότερης εξάπλωσης.



Σχ.8 Επιρροή βαθμού τοπολογίας

Επίσης παρατηρείται ότι για την ίδια τιμή μέσου βαθμού(ίσο με 4)σε δομημένα και μη p2p συστήματα,η αναλογία μόλυνσης αυξάνεται στα μη δομημένα.Αυτό συμβαίνει λόγω της ιδιότητας power-law των μη δομημένων p2p συστημάτων όπου για μεγάλες τιμές βαθμού,τα τερματικά έχουν πολύ μεγαλύτερη

συνδετικότητα.

Τέλος για τα μη δομημένα p2p συστήματα με ίδιο μέσο βαθμό (ω),καλύτερη επίδοση επιτυγχάνεται από αυτό που έχει τον χαμηλότερο βαθμό power-law(σ).Ο λόγος που συμβαίνει αυτό,είναι επειδή η power-law κατανομή έχει ιδιότητα long tail και έτσι μικρότερη τιμή του σ σημαίνει ότι η πιθανότητα ένα τερματικό να έχει k γείτονες (P(k)) είναι υψηλή.Συνεπώς λόγω αυξημένης συνδετικότητας,η αναλογία μόλυνσης είναι υψηλή όταν το σ είναι μικρό.

4.2.5 Σύνοψη:

- Yu,Chellappan,Wang,Xuan (2008 [32])
- Δύο στάδια στην επίθεση ενός ενεργού σκουληκιού.Η σάρωση του δικτύου για εύρεση πιθανών στόχων (τερματικών) και η μόλυνση αυτών αφού διαπιστωθεί ότι είναι ευάλωτοι.
- Offline p2p based επίθεση:

  α) ο επιτιθέμενος συλλέγει πληροφορίες για IP διευθύνσεις από το p2p σύστημα ενώ είναι εκτός σύνδεσης(offline)-σύνθεση hitlist σκουληκιού.

  β) Κατά την φάση επίθεσης όλα τα αρχικά μολυσματικά τερματικά εξαπολύουν επίθεσεις στα τερματικά των οποίων οι IP διευθύνσεις βρίσκονται στην hitlist(τερματικά στο super-p2p σύστημα).

  γ) Κατά τη δεύτερη φάση όλα τα μολυσματικά τερματικά συνεχίζουν τις επιθέσεις στο Internet (τερματικά στο non-p2p σύστημα) χρησιμοποιώντας μηχανισμό τυχαίας σάρωσης.

  Θεώρημα:

$$E^s(i+1)=\begin{cases} N^s(i)[1-(1-\frac{1}{m*u})^{S*M^s(i)}] & , M^s(i)\le m*u*P_3 \\ 0 & , M^s(i)>m*u*P_3 \end{cases}$$

τερματικά που μόλις μολύνθηκαν

όπου $M^s(0)=\mathrm{M}_0, N^s(0)=m*u*P_3$

$$E^n(i+1)=\begin{cases} 0 & , M^s(i)\le m*u*P_3 \\ N^n(i)[1-(1-\frac{1}{T})^{(S*M^n(i)+S*M^s(K)}] & , M^s(i)>m*u*P_3 \end{cases}$$

τερματικά που μόλις μολύνθηκαν

όπου $M^n(0)=0, N^n(K)=T*P_1*P_2-m*u*P_3, M^n(K)=M^s(K-1)$

και $K=\min(i)\ ,\forall i$ που ικανοποιεί την $M^s(i)>m*u*P_3$

- Online p2p based επίθεση:

α) Μόλις ένα μολυσματικό τερματικό εισέλθει στο p2p σύστημα,εξαπολύει επίθεση με υψηλή προτεραιότητα στους p2p γείτονές του.

β) Όταν δεν υπάρχει άλλο γειτονικό p2p τερματικό να σαρώσει, επιτίθεται με το 100% των δυνατοτήτων του στο Internet χρησιμοποιώντας μηχανισμό τυχαίας σάρωσης.

γ) Διαφορετική προσέγγιση για δομημένα p2p συστήματα όπου ο βαθμός τοπολογίας είναι σταθερός θ και μοντελοποιείται βάσει της κανονικής κατανομής και για μη δομημένα όπου ο βαθμός τοπολογίας ακολουθεί power-law.

Θεώρημα:

$$E^s(i+1) = N^s(i)\left[1-\left(1-\frac{1}{m*u}\right)^{\left(\left(\sum_{j=1}^{E^s(i)}\min(r_j,S)\right)+\frac{m*u}{T}*S*M^n(i)\right)}\right]$$

τερματικά που μόλις μολύνθηκαν

όπου r_j είναι ο βαθμός τοπολογίας του τερματικού j

και $M^s(0) = \mathrm{M}_0, N^s(0) = m*u*P_3$

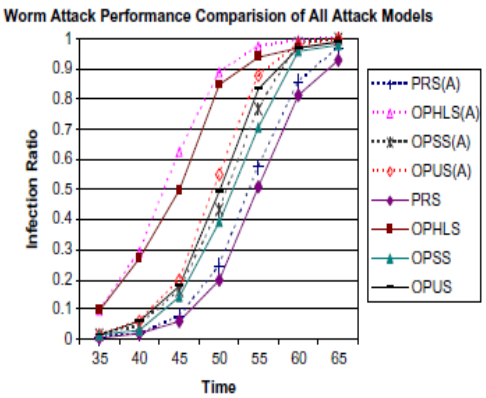
$$E^n(i+1) = N^n(i)\left[1-\left(1-\frac{1}{T}\right)^{\left(S*(M^n(i)+M^s(i))-\sum_{j=1}^{E^s(i)}\min(r_j,S)-\frac{m*u}{T}*S*M^n(i)\right)}\right]$$

τερματικά που μόλις μολύνθηκαν

όπου r_j είναι ο βαθμός τοπολογίας του τερματικού j

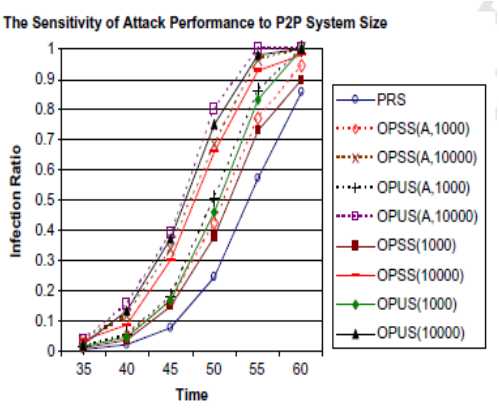
και $M^n(0) = 0, N^n(0) = T*P_1*P_2 - m*u*P_3$

- θ ↑ ή ω ↑ => r_j ↑ => $E^s(i)$ ↑ και $E^n(i)$ ↓ (η αύξηση του $E^s(i)$ είναι μεγαλύτερη από τη μείωση του $E^n(i)$) => $E(i)=E^s(i)+E^n(i)$ ↑ => η εξάπλωση του σκουληκιού επιταχύνεται.



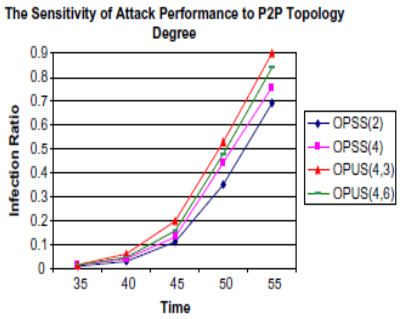
- Τα p2p based μοντέλα έχουν καλύτερη επίδοση από το PRS μοντέλο όπως επίσης το μοντέλο

OPHLS πετυχαίνει γρηγορότερη εξάπλωση σκουληκιού συγκρινόμενο με τα online-based μοντέλα επίθεσης.



- Όταν το μέγεθος του p2p συστήματος αυξάνεται, η επίδοση όλων των μοντέλων βελτιώνεται.Αυτό συμβαίνει γιατί η πιθανότητα, κατά τη διάρκεια μίας σάρωσης, να βρεθεί ευάλωτο τερματικό αυξάνεται και συνεπώς αυξάνεται και ο αριθμός των μολύνσεων.

- α) Στο δομημένο p2p σύστημα μία αύξηση στον βαθμό τοπολογίας, επιφέρει γρηγορότερη εξάπλωση σκουληκιού. Αυτό συμβαίνει γιατί όταν ο βαθμός αυξάνεται,αυξάνεται και η συνδετικότητα γεγονός που επιτρέπει στο σκουλήκι την σάρωση περισσότερων τερματικών και συνεπώς γρηγορότερης εξάπλωσης.



β) Για την ίδια τιμή μέσου βαθμού(ίσο με 4)σε δομημένα και μη p2p συστήματα,η αναλογία μόλυνσης αυξάνεται στα μη δομημένα.Αυτό συμβαίνει λόγω της ιδιότητας power-law των μη δομημένων p2p συστημάτων όπου για μεγάλες τιμές βαθμού,τα τερματικά έχουν πολύ μεγαλύτερη συνδετικότητα.

γ) Για τα μη δομημένα p2p συστήματα με ίδιο μέσο βαθμό (ω),καλύτερη επίδοση επιτυγχάνεται από αυτό που έχει τον χαμηλότερο βαθμό power-law(σ).Ο λόγος που συμβαίνει αυτό,είναι επειδή η power-law κατανομή έχει ιδιότητα long tail και έτσι μικρότερη τιμή του σ σημαίνει ότι η πιθανότητα ένα τερματικό να έχει k γείτονες (P(k)) είναι υψηλή.Συνεπώς λόγω αυξημένης συνδετικότητας,η αναλογία μόλυνσης είναι υψηλή όταν το σ είναι μικρό.

- Στις προσομοιώσεις θεωρείται ότι τα p2p τερματικά έχουν ίδια ευπάθεια με τα υπόλοιπα τερματικά του Internet πράγμα που δεν είναι ρεαλιστικό αφού τα p2p τερματικά στην πραγματικότητα είναι πιο ευπαθή.Συνεπώς,τα αποτελέσματα των προσομοιώσεων θεωρούνται αρκετά αισιόδοξα.

4.3 Μοντέλα εξάπλωσης Bluetooth σκουληκιών:

Την τελευταία δεκαετία όλο και περισσότερες ασύρματες κινητές συσκευές έκαναν την εμφάνιση τους και απέκτησαν τεράστια δημοτικότητα λόγω των διευκολύνσεων που προσέφεραν.Κύρια παραδείγματα αυτών,τα κινητά τηλέφωνα και τα PDAs.Είναι ευνόητο ότι όσο περισσότερο χρησιμοποιούνται οι συγκεκριμένες συσκευές τόσο περισσότεροι κακόβουλοι χρήστες θα σκοπεύουν στην επίθεση αυτών.Μέχρι τώρα έχουν βρεθεί πάνω από εκατό περιπτώσεις κακόβουλου λογισμικού (ιοί και σκουλήκια) που διαδίδονται μέσω διάφορων κινητών συσκευών. Αξίζει να αναφερθεί ότι κοινό χαρακτηριστικό των περισσότερων ιών και σκουληκιών που αναφέρθηκαν παραπάνω είναι ότι εκμεταλλεύονται τις δυνατότητες του Bluetooth για την εξάπλωσή τους.Το Bluetooth είναι μία στενού εύρους ραδιο τεχνολογία που σκοπεύει στην σύνδεση διαφορετικών ασύρματων συσκευών μεταξύ τους καταναλώνοντας ελάχιστη ενέργεια και έχοντας αρκετά χαμηλό κόστος.Μερικές από τις εφαρμογές του Bluetooth είναι τα ασύρματα ακουστικά,η dial-up δικτύωση και ο p2p διαμοιρασμός αρχείων.Παρακάτω θα παρουσιάσουμε μερικά γενικά χαρακτηριστικά της τεχνολογίας Bluetooth,θα διασαφηνίσουμε τη συμπεριφορά των Bluetooth σκουληκιών και θα συνεχίσουμε με την παρουσίαση του μοντέλου εξάπλωσης[34].

4.3.1 Γενικά χαρακτηριστικά της τεχνολογίας Bluetooth:

Όπως αναφέραμε και προηγουμένως,το Bluetooth είναι μία στενού εύρους ραδιο τεχνολογία που σκοπεύει στην σύνδεση διαφορετικών ασύρματων συσκευών μεταξύ τους.Λειτουργεί στην μπάντα συχνοτήτων 2.4- Ghz και τα κανάλια του μοιράζονται στις συσκευές μέσω τεχνικής TDD (Time Division Duplexing).Επίσης χρησιμοποιεί διασπορα φάσματος με μεταπήδηση συχνότητας (frequency-hopping

spread spectrum) για μείωση στο ελάχιστο των παρεμβολών από γειτονικές συσκευές.Μία συσκευή Bluetooth μπορεί να λειτουργήσει σε μία εκ των τριών ενεργειακών κλάσεων 1,2 ή 3 που αντιστοιχούν σε εμβέλεια 100,10 και 0.1 m.

Όταν μία συσκευή Bluetooth θέλει να εντοπίσει άλλες συσκευές που βρίσκονται στην κοντινή της περιοχή,στέλνει πακέτα εύρεσης (inquiry packets) σε μία ακολουθία συχνοτήτων, αλλάζοντας συχνότητες 3200 φορές το δευτερόλεπτο,γεγονός που επιτρέπει την κάλυψη του φάσματος συχνοτήτων πολύ γρήγορα.Από την άλλη πλευρά, η συσκευή που βρίσκεται σε ανιχνεύσιμη κατασταση αλλάζει συχνότητες μία φορά κάθε 1.28 δευτερόλεπτα.Η διαφορά αυτή στην ταχύτητα αλλαγής συχνοτήτων διαβεβαιώνει ότι κάποια στιγμή η ανιχνεύσιμη συσκευή θα βρεθεί στην ίδια συχνότητα με την συσκευή που στέλνει τα πακέτα.Όταν η ανιχνεύσιμη συσκευή λάβει ένα πακέτο εύρεσης απαντάει στη συσκευή που κάνει την ανίχνευση,με ένα πακέτο στο οποίο περιέχεται η διεύθυνσης της και πληροφορίες συγχρονισμού και αναμένει μήνυμα εντοπισμού(page)[34].

Από τη στιγμή που η συσκευή έχει ανακαλύψει όλες τις γειτονικές της συσκευές,είναι πιθανό να θέλει να συνδεθεί με μία ή περισσότερες από αυτές.Η Bluetooth σύνδεση δύο γειτονικών συσκευών επιτυγχάνεται μέσω της διαδικασίας εντοπισμού(paging process).Η συγκεκριμένη διαδικασία είναι παρόμοια με τη διαδικασία εύρεσης(inquiry process)μόνο που στην διαδικασία εντοπισμού,η συσκευή που κάνει τον εντοπισμό καθορίζει σαφώς την διευθυνση της συσκευής με την οποία θέλει να συνδεθεί.Από τη στιγμή που εδραιωθεί η σύνδεση,η συσκευή που έκανε τον εντοπισμό ονομάζεται master και η συσκευή που εντοπίστηκε slave.Μία master συσκευή μπορεί να έχει μέχρι και επτά slave συσκευές σε ένα piconet(βασική δομική μονάδα ενός δικτύου Bluetooth —πολλά piconets δημιουργούν ένα scatternet).Στο piconet οι slave συσκευές μπορούν να επικοινωνούν μόνο με την master συσκευή,και όχι μεταξύ τους.Η master συσκευή επίσης καθορίζει το πως διαμοιράζεται το εύρος ζώνης του καναλιού στους slaves.

Μία σύνδεση Bluetooth έχει μέγιστη χωρητικότητα 1Mbps και υπάρχουν δύο κατηγορίες συνδέσεων Bluetooth.Οι SCO (Synchronous connection-oriented) για τις επικοινωνίες φωνής και οι ACL(Asynchronous connectionless) για επικοινωνίες δεδομένων.

Παρακάτω και μέχρι το τέλος του υποκεφαλαίου,θα θεωρούμε ότι τα Bluetooth σκουλήκια χρησιμοποιούν ACL συνδέσεις για την εξάπλωσή τους.

4.3.2 Bluetooth σκουλήκια:

Όταν ένα Bluetooth σκουλήκι ενεργοποιηθεί,αρχίζει και ψάχνει για συσκευές με ενεργοποιημένο Bluetooth που βρίσκονται εντός εμβέλειας της ήδη μολυσμένης συσκευής.Σε αυτή τη φάση η μολυσμένη συσκευή στέλνει πακέτα εύρεσης και περιμένει απαντήσεις.Επειδή δεν υπάρχει συγκεκριμένος αριθμός απαντήσεων που πρέπει να ληφθούν,το σκουλήκι καθορίζει τον μέγιστο αριθμό απαντήσεων N_{inq}^{t0} καθώς και τον μέγιστο χρόνο που θέλει να περιμένει για τη λήψη απαντήσεων T_{inq}^{t0}. Άν ληφθούν N_{inq}^{t0} απαντήσεις πριν ολοκληρωθεί ο χρόνος T_{inq}^{t0} τότε με την άφιξη της N_{inq}^{t0} -οστης απάντησης το σκουλήκι σταματάει τη φάση εύρεσης και προχωράει στην επόμενη φάση.Διαφορετικά αν ολοκληρωθεί ο χρόνος T_{inq}^{t0} το σκουλήκι σταματάει τη φάση της εύρεσης ανεξαρτήτως του αριθμού των απαντήσεων που έχουν ληφθεί.Αφού λοιπόν συλλεχθεί η λίστα με τις παραπάνω συσκευές,επιχειρεί τις παρακάτω διαδικασίες σε κάθε συσκευή ξεχωριστά.(1)εδραίωση σύνδεσης με αυτή, (2)εκτίμηση δυνατότητας μόλυνσης της,(3)αντιγραφή κακόβουλου κώδικα στη συσκευή του θύματος και (4) αποδύνδεση από αυτή.

Λόγω αστάθειας των συνδέσεων του κινητού δικτύου,οποιαδήποτε από τις παραπάνω διαδικασίες είναι πιθανό να αποτύχει.Γι'αυτό έχει οριστεί χρονικός περιορισμός για την ολοκλήρωση κάθε μίας,έτσι ώστε να ανιχνεύονται πιθανά σφαλματα σύνδεσης.

Στο συγκεκριμένο μοντέλο εξάπλωσης Bluetooth σκουληκιών που προτάθηκε λαμβάνεται ως δεδομένο ότι κατά τη διαδικασία εκτίμησης της δυνατότητας μόλυνσης,η γειτονική συσκευή μπορεί να στείλει μία από τις παρακάτω απαντήσεις:

α) απάντηση REJECTED:που υποδεικνύει ότι η συσκευή είναι μη ευάλωτη.

β) απάντηση UNINFECTED:που υποδεικνύει ότι η συσκευή είναι ευάλωτη και μη μολυσμένη.

γ) απάντηση INFECTED:που υποδεικνύει ότι η συσκευή είναι ευάλωτη αλλά μολυσμένη.

Τέλος όταν όλες οι γειτονικές συσκευές της λίστας ελεγχθούν ή μολυνθούν το σκουλήκι παραμένει ανενεργό για συγκεκριμένο χρονικό διάστημα,με το πέρας του οποίου,μπαίνει σε έναν καινούριο κύκλο μόλυνσης και επαναλαμβάνει τις παραπάνω διαδικασίες.

4.3.3 Μεθοδολογία μοντέλου:

Το μοντέλο που προτείνεται για την εξάπλωση των Bluetooth σκουληκιών είναι ντετερμινιστικό και διακριτού χρόνου.Έστω i(t) η μέση πυκνότητα μολυσματικών συσκευών στο δίκτυο,δοθέντος χρόνου t και ότι το σκουλήκι ξεκινάει την διαδικασία διάδοσης του σε χρόνο t_0 με αρχική πυκνότητα μόλυνσης $i(t_0)$.Δεδομένης της γνώσης της προόδου εξάπλωσης του σκουληκιού $i(t_k)$ σε χρόνο t_k όπου k>=0 , το μοντέλο καθορίζει την επόμενη χρονική στιγμή t_{k+1} και την κατάσταση της εξάπλωσης του σκουληκιού $i(t_{k+1})$. Με $T_{cycle}(t)$ παριστάνεται η διάρκεια ενός κύκλου μόλυνσης που ξεκινάει τη χρονική στιγμή t και επιλέγεται σαν βήμα χρόνου $t_{k+1} - t_k = T_{cycle}(t_k)$. Επιπλέον ανάμεσα σε δύο οποιεσδήποτε χρονικές στιγμές t_k και t_{k+1} χρησιμοποιείται η ακόλουθη εξίσωση για την εκτίμηση της καμπύλης εξάπλωσης του σκουληκιού(1):

$$\frac{di(t)}{dt} = \beta(t)i(t)(\rho(t) - i(t))$$

όπου ρ(t) είναι η μέση πυκνότητα συσκευών και β(t) ο ρυθμός μόλυνσης σε χρόνο t.

Για να προκύψουν τα $T_{cycle}(t)$ και β(t) γίνονται οι εξής υποθέσεις:

α) όλες οι συσκευές είναι ομογενώς ανακατεμένες

β) η συμπεριφορά μίας μολυσματικής συσκευής σε χρόνο t είναι μία ντετερμινιστική συνάρτηση της πυκνότητας συσκευών ρ(t),της προόδου εξάπλωσης του σκουληκιού i(t) και των στατιστικών ιδιοτήτων της κινητικότητας των συσκευών.

γ) όλες οι μολυσματικές συσκευές δεδομένου χρόνου t έχουν παρόμοιο κύκλο μόλυνσης αλλά μπορούν να βρίσκονται σε διαφορετικές φάσεις στον κύκλο μόλυνσης.

Παρακάτω θα αναλυθεί ένας κύκλος μόλυνσης από όπου θα προκύψει η διάρκεια του ($T_{cycle}(t)$) και το πλήθος των καινούριων μολύνσεων στον συγκεκριμένο

κύκλο (α(t)) και θα συζητηθεί το πώς προκύπτει ο ρυθμός β(t) από το α(t) καθώς επίσης θα χρησιμοποιηθεί η εξίσωση (1) για την εκτίμηση της καμπύλης εξάπλωσης του σκουληκιού.

Έστω μία συσκευή 0 , χωρίς βλάβη της γενικότητας , ξεκινάει τη διαδικασία εύρεσης σε χρόνο t,και με $T_{inq}(t)$ παριστάνεται η μέση διάρκεια της φάσης εύρεσης,δεδομένου χρόνου t.Οι γείτονες μίας συσκευής με ενεργοποιημένο Bluetooth (στην περιπτωση μας,της συσκευής 0) διαχωρίζονται σε δύο κατηγορίες.Στην πρώτη κατηγορία ανήκουν οι στιγμιαίοι γείτονες(instantaneous),οι οποίοι είναι αυτοί που βρίσκονται εντός εμβέλειας της συσκευής 0 όταν αυτή ξεκινάει τη διαδικασία εύρεσης.Το μέσο πλήθος τους σε χρόνο t είναι στην ουσία $J_{in}(t)$,δηλαδή ο μέσος βαθμός κόμβου σε χρόνο t.Με την πάροδο του χρόνου,κάποιοι στιγμιαίοι γείτονες μετακινούνται εκτός εμβέλειας της συσκευής 0 και την ίδια στιγμή πιθανόν κάποιοι καινούριοι εισέρχονται στην εμβέλεια της.Αυτοί οι καινούριοι γείτονες αποτελούν την δεύτερη κατηγορία που είναι οι απρόοπτοι γείτονες (contingent) της συσκευής 0 και παριστάνονται με $J_{co}(t)$.Προφανώς,το πλήθος τους εξαρτάται από τη διάρκεια της φάσης εύρεσης.Η διαδικασία άφιξης καινούριων γειτόνων είναι διαδικασία Poisson με μέσο ρυθμό άφιξης $\lambda_{ne}(t)$.Χρησιμοποιώντας την ιδιότητα της Poisson διαδικασίας,PASTA (Poisson Arrivals See Time Averages),το πλήθος των γειτόνων που η συσκευή 0 συναντά κατά τη φάση εύρεσης είναι $H_{inq}(t)=J_{in}(t)+J_{co}(t)$,όπου $J_{co}(t)=\lambda_{ne}(t)T_{inq}(t)$.

Όπως αναφέραμε προηγουμένως,θεωρώντας τον πληθυσμό των συσκευών ομογενώς ανακατεμένο στο δίκτυο,το μοντέλο εξάπλωσης του σκουληκιού μπορει να περιγραφεί από την εξίσωση (1):

$$\frac{di(t)}{dt} = \beta(t)i(t)(\rho(t) - i(t))$$

Στη συνέχεια χρησιμοποιώντας την παραπάνω εξίσωση καθώς και το πλήθος καινούριων μολύνσεων α(t) σε διάρκεια χρόνου ενός κύκλου μόλυνσης $T_{cycle}(t)$ είναι δυνατή η εκτίμηση του ρυθμού μόλυνσης β(t) από την παρακάτω εξίσωση(2):

$$\frac{di(t)}{dt} = \beta(t)i(t)(\rho(t) - i(t)) = \frac{a(t)}{T_{cycle}(t)}i(t) \Rightarrow$$

$$\beta(t) = \frac{a(t)}{(\rho(t) - i(t))T_{cycle}(t)}$$

Πριν προχωρήσουμε παρακάτω είναι απαραίτητο να προσδιοριστεί ο τρόπος υπολογισμού του πλήθους των καινούριων μολύνσεων α(t).

Υπολογισμός του α(t):

Αρχικά ορίζουμε τον αριθμό των γειτόνων,της συσκευής 0,που ανακαλύφθηκαν κατά τη φάση εύρεσης ως R(t) με

$$R(t) = \min\{N_{inq}^{to}, N_{rsp}^{ins}(t) + N_{rsp}^{co}(t)\}$$

όπου N_{inq}^{t0} είναι ο μέγιστος αριθμός αναμενόμενων απαντήσεων κατά τη φάση εύρεσης,όπως αναφέρθηκε προηγουμένως,και N_{rsp}^{ins} και N_{rsp}^{co} το μέσο πλήθος στιγμιαίων και απρόοπτων γειτόνων ,αντίστοιχα,που ανακαλύπτονται από τη συσκευή 0.

Στη συνέχεια ορίζουμε αναδρομικά μία συνάρτηση Ω στην οποία βασίζεται ο υπολογισμός του α(t).

Η Ω λοιπόν ορίζεται ως εξής:

$$\Omega(t, k, \tau_s^{(k)}(t), \vec{V}) = \begin{cases} 0 & , k > R(t) \\ \omega & , k \le R(t) \end{cases}$$

όπου k είναι ο βαθμός της συσκευής,$\tau_s^{(k)}$ είναι η διάρκεια της περιόδου που ξεκινάει με την έναρξη της φάσης εύρεσης της συσκευής 0 και τελειώνει με την έναρξη της διαδικασίας επεξεργασίας του γείτονα k και **V** ένα διάνυσμα πέντε στοιχείων.

Επιπλέον το ω ορίζεται ως εξής:

$$\omega = P_{conn}^{fail}(t,\tau_s^{(k)}(t))(\vec{V}[1]+\Omega(t,k+1,\tau_s^{(k)}(t)+T_{conn}^{to}))+$$
$$P_{prb}^{fail}(t,\tau_s^{(k)}(t))(\vec{V}[2]+\Omega(t,k+1,\tau_s^{(k)}(t)+T_{conn}^{good}(t)+T_{prb}^{to}))+$$
$$P_{\inf}^{prb}(t,\tau_s^{(k)}(t))(\vec{V}[3]+\Omega(t,k+1,\tau_s^{(k)}(t)+T_{conn}^{good}(t)+\frac{S_{prb}}{\eta(t)}))+$$
$$P_{rep}^{succ}(t,\tau_s^{(k)}(t))(\vec{V}[4]+\Omega(t,k+1,\tau_s^{(k)}(t)+T_{conn}^{good}(t)+\frac{S_{prb}}{\eta(t)}+\frac{S_{worm}}{\eta(t)}))+$$
$$P_{rep}^{fail}(t,\tau_s^{(k)}(t))(\vec{V}[5]+\Omega(t,k+1,\tau_s^{(k)}(t)+T_{conn}^{good}(t)+\frac{S_{prb}}{\eta(t)}+T_{rep}^{to}))$$

όπου

T_{conn}^{to}	τιμή timeout εδραίωσης σύνδεσης
T_{prb}^{to}	τιμή timeout διαδικασίας εξερεύνησης
T_{rep}^{to}	τιμή timeout αντιγραφής σκουληκιού
$T_{conn}^{good}(t)$	μεσος χρονος για επιτυχη εδραιωση συνδεσης
P_{conn}^{fail}	πιθανότητα αποτυχίας εδραίωσης σύνδεσης
P_{prb}^{fail}	πιθανότητα αποτυχίας εξερεύνησης της μολυσματικής κατάστασης ενός γείτονα
$P_{\inf}^{prb}$	πιθανότητα να βρεθεί ότι ο γείτονας που εξερευνάται είναι ήδη μολυσμένος
P_{rep}^{succ}	πιθανότητα επιτυχούς αντιγραφής του κώδικα του σκουληκιού στο θύμα
P_{rep}^{fail}	πιθανότητα ανεπιτυχούς αντιγραφής του κώδικα του σκουληκιού στο θύμα
S_{prb}	μέγεθος πακέτου εξερεύνησης
S_{worm}	μέγεθος κώδικα σκουληκιού
$\eta(t)$	ρυθμός μετάδοσης των δεδομένων

και τελικά **$\alpha(t)=\Omega(t,1,T_{inq}(t),<0,0,0,1,0>)$**.

Εν συνεχεία λύνοντας την εξίσωση (1) προκύπτει ότι η εξάπλωση του σκουληκιού περιγράφεται από την σχέση(3):

$$i(t+\Delta t)=\frac{i(t)\rho(t)}{i(t)+(\rho(t)-i(t))e^{-\beta(t)\rho(t)\Delta t}}$$

και συνεπώς η καινουρια πυκνότητα μολυσματικών συσκευών μετά την ολοκλήρωση ενός κύκλου μόλυνσης είναι(4):

$$i(t+T_{cycle}(t)) = \frac{i(t)\rho(t)}{i(t)+(\rho(t)-i(t))e^{-a(t)\rho(t)/(\rho(t)-i(t))}}$$

Η παραπάνω εξίσωση οδηγεί σε μία μέθοδο υπολογισμού ολόκληρης της καμπύλης μόλυνσης.

Έστω t_0=0 και έστω ότι σε χρόνο t_0 υπάρχει μόνο μία μολυσματική συσκευή.Άρα $i(t_0)=\rho(t_0)/N_{dev}(t_0)$ όπου $N_{dev}(t)$ είναι το συνολικό πλήθος των συσκευών σε χρόνο t.Ξεκινώντας από τη χρονική στιγμή t_0 υπολογίζουμε το $T_{cycle}(t_k)$ και στη συνέχεια ορίζουμε αναδρομικά τα t_{k+1} και $i(t_{k+1})$ ακολούθως(5):

$$t_{k+1} = t_k + T_{cycle}(t_k)$$
$$i(t_{k+1}) = \frac{i(t_k)\rho(t_k)}{i(t_k)+(\rho(t_k)-i(t_k))e^{-a(t_k)\rho(t_k)/(\rho(t_k)-i(t_k))}}$$

Παρ'όλα αυτά παρατηρείται ότι υπάρχουν μερικά προβλήματα με την παραπάνω προσέγγιση.Πρώτον κατά την πρώιμη φάση εξάπλωσης του σκουληκιού,οι μολυσματικές συσκευές είναι συσσωρευμένες η μία κοντά στην άλλη γιατί απαιτείται κάποιος χρόνος για την εξάπλωση τους στην περιοχή.Μία παραδοχή του συγκεκριμένου μοντέλου είναι ότι οι μολυσματικές και μη συσκευές είναι ομογενώς ανακατεμένες στο δίκτυο.Αυτό το πρόβλημα εκδηλώνεται ειδικά όταν ένας μικρός αριθμός συσκευών είναι αραιά κατανεμημένος σε μία μεγάλη περιοχή.

Για τη διευθέτηση του προβλήματος αυτού,εισάγεται ένα κάτω όριο στην πυκνότητα των μολυσματικών συσκευών ως εξής:

Έστω μία μολυσματική συσκευή που ξεκινάει την φάση εύρεσης σε χρόνο t και κινείται κατά τη φάση αυτή,σε μία ευθεία γραμμή.Η περιοχή που καλύπτεται από το ραδιο-σήμα της συγκεκριμένης συσκευής κατά τη φάση αυτή είναι (6):

$$S_{inq} = \pi r_{ra}{}^2 + 2r_{ra}u(t)T_{inq}(t)$$

όπου u(t) είναι η μέση ταχύτητα της συσκευής,και r_{ra} η ραδιο-εμβέλεια της συσκευής. Στην περιοχή,που καλύπτεται από την μολυσματική συσκευή,υπάρχει μία τουλάχιστον μολυσματική συσκευή η οποία είναι αυτή η ίδια.

Ορίζουμε (7):

$$i'(t) = \max\{i(t), \frac{1}{S_{inq}(t)}\}$$

Στην συνέχεια υπολογίζουμε τα $T_{cycle}(t_k)$ και $\alpha(t_k)$ χρησιμοποιώντας το i'(t) αντί του i(t) και έτσι η εξίσωση (5) γίνεται:

$$i(t_{k+1}) = \frac{i'(t_k)\rho(t_k)}{i'(t_k) + (\rho(t_k) - i'(t_k))e^{-a(t_k)\rho(t_k)/(\rho(t_k)-i'(t_k))}}$$

Το δεύτερο πρόβλημα με τις σχεσεις (5) σχετίζεται με την παραδοχή ότι οι καινούριες μολύνσεις σε έναν κύκλο μόλυνσης είναι ισοκατανεμημένες στον κύκλο μόλυνσης.Αν οι καινούριες μολύνσεις είναι πολλές η παραδοχή είναι λογική,γιατί η φάση μίας μολυσματικής συσκευής στον κύκλο μόλυνσης θεωρείται τυχαία.Παρ'όλα αυτά,στην πρώιμη φάση της μόλυνσης,μία συσκευή που μόλις μολύνθηκε μπαίνει κατ'ευθείαν σε κατάσταση ενεργού σκαναρίσματος.Άρα χρησιμοποιώντας τη σχέση (5) για την πρόβλεψη της εξάπλωσης ενός σκουληκιού υποτιμάται η ταχύτητα εξάπλωσης του.Επιπλέον αν το $\beta(t_k)$ είναι μεγαλύτερο,τότε υπάρχουν περισσότερες νέες μολύνσεις σε έναν κύκλο μόλυνσης και άρα η εκτίμηση λάθους είναι μεγαλύτερη.Συνεπώς μειώνουμε το $T_{cycle}(t_k)$ που βασίζεται στο $\beta(t_k)$ στις πρώτες λίγες επαναλήψεις και έτσι το μοντέλο υπολογισμού του $T_{cycle}(t_k)$ έχει ως εξής:

$$T_{cycle}(t_k) = \begin{cases} T_{inq}(t_k) + T_{proc}(t_k) + e^{-2\beta(t_k)T_{idle}^{to}} & ,\alpha\nu\ k < 3 \\ T_{inq}(t_k) + T_{proc}(t_k) + T_{idle}^{to} & ,\alpha\nu\ k \geq 3 \end{cases}$$

όπου $T_{proc}(t)$ ο συνολικός χρόνος που ξοδεύει μία συσκευή στην επεξεργασία των συσκευών που ανακάλυψε και T_{idle}^{t0} η διάρκεια της φάσης αναμονής.

Το τρίτο πρόβλημα με το μοντέλο είναι ότι υπολογίζει τον ρυθμό αύξησης των σκουληκιών βασισμένο στην κατάσταση μόλυνσης τη χρονική στιγμή t_k και έτσι υποθέτει ότι αυτός ο ρυθμός αύξησης παραμένει σταθερός κατά τη διάρκεια ενός κύκλου μόλυνσης,που ξεκίνησε τη χρονική στιγμή t και τελείωσε τη χρονική στιγμή

t_{k+1}.Για τις μολυσματικές συσκευές που ξεκινούν τον κύκλο μόλυνσης τους μετά τη χρονική στιγμή t αλλά πριν την t_{k+1} το α είναι υπερεκτιμημένο.Για να ξεπεραστεί το συγκεκριμένο πρόβλημα επαναπροσδιορίζουμε τον υπολογισμό του $i(t_{k+1})$ ως εξής: Αρχικά υπολογίζουμε το $\alpha(t_k)$ όπως προηγουμένως και στη συνέχεια εκτιμούμε την πυκνότητα των μολυσματικών συσκευών για χρόνο t_x ,όπου $t_x = t_k + T_{cycle}(t_k) - T_{proc}(t_k)$. Ουσιαστικά, t_x είναι η αργότερη χρονική στιγμή που μία μολυσματική συσκευή τελειώνει την φάση εύρεσης έτσι ώστε να μπορεί να τελειώσει πριν από χρόνο $t_k + T_{cycle}(t_k)$ την επεξεργασία όλων των γειτόνων που ανακαλύφθηκαν.
Η εκτιμώμενη κατάσταση μόλυνσης για χρόνο t_x είναι

$$i(t_x) = \frac{i(t_k)\rho(t_k)}{i'(t_k) + (\rho(t_k) - i'(t_k))e^{\frac{-a(t_k)\rho(t_k)}{\rho(t_k)-i'(t_k)} * \frac{t_x - t_k}{T_{cycle}(t_k)}}}$$

Βασιζόμενοι στην παραπάνω σχέση,είναι δυνατός ο υπολογισμός του $\alpha(t_x)$.Ορίζουμε την παρακάτω σχέση:

$$a' = \frac{\rho(t_k) - i(t_k)}{\rho(t_k)} a(t_k) + \frac{i(t_k)}{\rho(t_k)} a(t_x)$$

Συνεπώς η καινούρια εξίσωση για τον υπολογισμό του $i(t_{k+1})$ είναι:

$$i(t_{k+1}) = \frac{i(t_k)\rho(t_k)}{i'(t_k) + (\rho(t_k) - i'(t_k))e^{-a'\rho(t_k)/(\rho(t_k)-i'(t_k))}}$$

4.3.4 Προσομοίωση-συμπεράσματα:

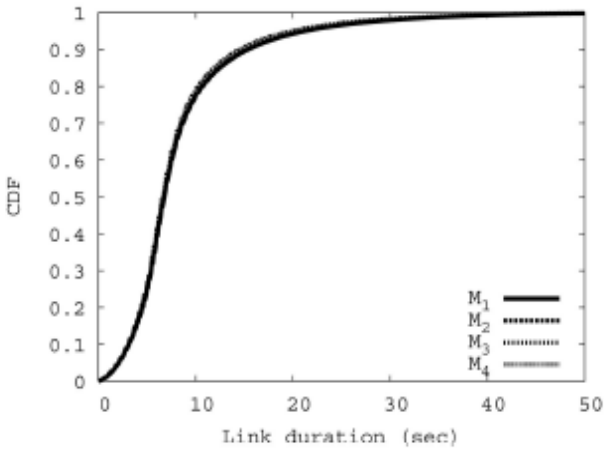
Το σύστημα εξισώσεων που προτάθηκε παραπάνω για την περιγραφή του μοντέλου εξάπλωσης ενός Bluetooth σκουληκιού είναι αρκετά δύσκολο να λυθεί αναλυτικά,λόγω της μεγάλης πολυπλοκότητας του.Συνεπώς κρίνεται απαραίτητη η χρήση λογισμικού για την εύρεση αριθμητικής λύσης.Στην συγκεκριμένη περίπτωση χρησιμοποιήθηκε το λογισμικό Octave[33]απ'όπου προέκυψαν και οι αριθμητικές λύσεις του συστήματος εξισώσεων που προτάθηκε.Στη συνέχεια χρησιμοποιήθηκε το

λογισμικό προσομοίωσης δικτύου ns-2 με το πακέτο προσομοίωσης UCBT Bluetooth.Για να εκτιμηθεί η ακρίβεια του προτεινόμενου μοντέλου,έγιναν πειράματα με διάφορες παραμέτρους Bluetooth και κινητικότητας.Σε όλες τις περιπτώσεις θεωρείται ότι μία Bluetooth συσκευή κινείται σε μία τετράγωνη περιοχή σύμφωνα με το μοντέλο τυχαίου περιπάτου,στην οποία περιοχή αλλάζει την κατεύθυνση και την ταχύτητα της(η συσκευή)κάθε 30 δευτερόλεπτα.Η ταχύτητα της συσκευής είναι ομοιόμορφα κατανεμημένη στις τιμές 1m/s και 2m/s[34].Ο παρακάτω πίνακας (Σχ.9) παρουσιάζει τις τιμές των παραμέτρων κινητικότητας που χρησιμοποιούνται στα πειράματα

ID	N_{dev}	S_{dev}	λ_{ne}	J_{in}
M_1	50	$75\times75\ m^2$	0.5239	2.4751
M_2	200	$75\times75\ m^2$	2.1199	10.0088
M_3	200	$150\times150\ m^2$	0.5753	2.6651
M_4	800	$150\times150\ m^2$	2.3089	10.6693

Σχ.9 Τιμές παραμέτρων κινητικότητας πειραμάτων

και η παρακάτω γραφική παράσταση(Σχ.10)παρουσιάζει την αθροιστική συνάρτηση κατανομής (CDF) της διάρκειας των συνδέσεων που αντιστοιχούν στα παραπάνω τέσσερα σενάρια κινητικότητας.



Σχ.10 Αθροιστική συνάρτηση κατανομής διάρκειας συνδέσεων

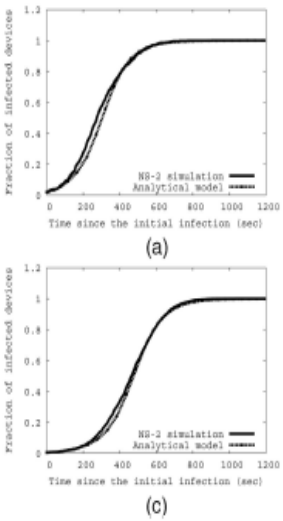
Παρόλο που σε καθένα από τα τέσσερα σενάρια υπάρχει διαφορετική πυκνότητα συσκευών ,οι CDFs των χρόνων των συνδέσεων τους είναι πολυ κοντά η μία στην άλλη.Αυτό συμβαίνει γιατι,η πυκνότητα των συσκευών επιρρεάζει το πόσο συχνά δύο Bluetooth συσκευές θα συναντηθούν.Όμως από τη στιγμή που η μία βρεθεί εντός εμβέλειας επικοινωνίας της άλλης τότε αυτό που καθορίζει την διάρκεια της σύνδεσης είναι οι παράμετροι κινητικότητας (παραπάνω πίνακας) που σχετίζονται με τον χρόνο που θα περάσει

μέχρι να βγει η μία συσκευή εκτός εμβέλειας της άλλης.

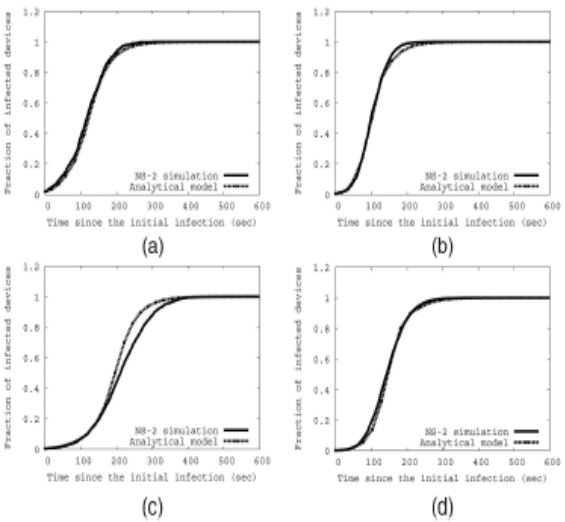
Επιπλέον χρησιμοποιούνται δύο ομάδες Bluetooth παραμέτρων W_1 και W_2.Στην W_1 έχουμε $T_{inq}^{t0}= 10.24\text{sec}$, $T_{idle}^{t0}= 20\text{sec}$ και $N_{inq}^{t0}= 5$.Ενώ στην W_2 έχουμε $T_{inq}^{t0}= 5.12\text{sec}$, $T_{idle}^{t0}= 10\text{sec}$ και $N_{inq}^{t0}= 3$.Συνεπώς προκύπτουν συνολικά οκτώ σενάρια.Για κάθε ένα από αυτά το ns-2 τρέχει 20 προσομοιώσεις στις οποίες η μολυσματική συσκευή που ξεκινάει τη μόλυνση επιλέγεται τυχαία από ολόκληρο τον πληθυσμό.Μερικές ακόμα παράμετροι που χρησιμοποιούνται στα πειράματα είναι οι εξής:

$r_{ra}= 10\text{m}$, $S_{worm}= 20{,}000\,bytes$, $S_{prb}= 27\,bytes$,

$T_{conn}^{t0}= 5.12\text{sec}$, $T_{prb}^{t0}= 1\text{sec}$, $T_{rep}^{t0}= 10\text{sec}$, $T_{disc}^{t0}= 0.1\text{sec}$.



Σχ.11:Καμπύλες μόλυνσης βάσει ομάδας παραμέτρων Bluetooth W_1.(a)σενάριο κινητικότητας M1.(b)σενάριο M2 (c)σενάριο M3 (d)σενάριο M4



Σχ.12:Καμπύλες μόλυνσης βάσει ομάδας παραμέτρων Bluetooth W_2.(a)σενάριο κινητικότητας M1.(b)σενάριο M2 (c)σενάριο M3 (d)σενάριο M4

Στα Σχ.11 και Σχ.12 παρουσιάζονται τα ποσοστά μολυσματικών συσκευών συναρτήσει του χρόνου εξάπλωσης του σκουληκιού,όπως προέκυψαν από το μοντέλο και από τον μέσο όρο 20 προσομοιώσεων για κάθε ένα σενάριο.Είναι εμφανές ότι οι καμπύλες που προέκυψαν από το μοντέλο ταιριάζουν αρκετά ,στις περισσότερες των περιπτώσεων, με τα αποτελέσματα των προσομοιώσεων.Μοναδική εξαίρεση αποτελούν τα αποτελέσματα βάσει σεναρίου κινητικότητας M_3 και Bluetooth παραμέτρων W_2 όπου το μοντέλο υπερεκτιμά την ταχύτητα εξάπλωσης του

σκουληκιού κατά την τελευταία φάση της εξάπλωσής του.

4.3.5 Σύνοψη:

- Yan, Eidenbenz (2009 [34])
- βασισμένο στο SI μοντέλο.
- Bluetooth σκουλήκια: Όταν ένα Bluetooth σκουλήκι ενεργοποιηθεί,αρχίζει και ψάχνει για συσκευές με ενεργοποιημένο Bluetooth που βρίσκονται εντός εμβέλειας της ήδη μολυσμένης συσκευής.Όταν συλλέξει τη λίστα με τις παραπάνω συσκευές,επιχειρεί τις παρακάτω διαδικασίες σε κάθε συσκευή ξεχωριστά.(1)εδραίωση σύνδεσης με αυτή, (2)εκτίμηση δυνατότητας μόλυνσης της,(3)αντιγραφή κακόβουλου κώδικα στη συσκευή και (4) αποδύνδεση από αυτή.-κύκλος μόλυνσης.
- Κατά τη διαδικασία εκτίμησης της δυνατότητας μόλυνσης,η γειτονική συσκευή μπορεί να ανταποκριθεί ως εξής:

  α) απάντηση REJECTED:που υποδεικνύει ότι η συσκευή είναι μη ευάλωτη.

  β) απάντηση UNINFECTED:που υποδεικνύει ότι η συσκευή είναι ευάλωτη και μη μολυσμένη.

  γ) απάντηση INFECTED:που υποδεικνύει ότι η συσκευή είναι ευάλωτη αλλά μολυσμένη.
- Έστω t_0=0 και έστω ότι σε χρόνο t_0 υπάρχει μόνο μία μολυσματική συσκευή.

$$t_{k+1} = t_k + T_{cycle}(t_k)$$

$$i(t_{k+1}) = \frac{i(t_k)\rho(t_k)}{i(t_k) + (\rho(t_k) - i(t_k))e^{-a(t_k)\rho(t_k)/(\rho(t_k)-i(t_k))}}$$

- Προβλήματα:

  α) όταν ένας μικρός αριθμός συσκευών είναι αραιά κατανεμημένος σε μία μεγάλη περιοχή(το μοντέλο θεωρεί ότι οι μολυσματικές και μη συσκευές είναι ομογενώς κατανεμημένες στο δίκτυο).

  Λύση: εισαγωγή κάτω ορίου στην πυκνότητα των μολυσματικών συσκευών.

β) το μοντέλο θεωρεί ότι οι καινούριες μολύνσεις σε έναν κύκλο μόλυνσης είναι ισοκατανεμημένες στον κύκλο μόλυνσης.Όμως στην πρώιμη φάση της μόλυνσης,μία συσκευή που μόλις μολύνθηκε μπαίνει κατ'ευθείαν σε κατάσταση ενεργού σκαναρίσματος και έτσι στη συγκεκριμένη φάση η ταχύτητα εξάπλωσης του σκουληκιού υποτιμάται.

Λύση:ο υπολογισμός του $T_{cycle}(t_k)$ μειώνεται κατάλληλα στις πρώτες λίγες επαναλήψεις.

γ) το μοντέλο είναι ότι υπολογίζει τον ρυθμό αύξησης τωνσκουληκιών βασισμένο στην κατάσταση μόλυνσης τη χρονική στιγμή t_k και έτσι υποθέτει ότι αυτός ο ρυθμός αύξησης παραμένει σταθερός κατά τη διάρκεια ενός κύκλου μόλυνσης,που ξεκίνησε τη χρονική στιγμή t και τελείωσε τη χρονική στιγμή t_{k+1}.Όμως, για τις μολυσματικές συσκευές που ξεκινούν τον κύκλο μόλυνσης τους μετά τη χρονική στιγμή t αλλά πριν την t_{k+1} το α είναι υπερεκτιμημένο.

Λύση:επανυπολογισμός του α.

- Τελικά,

$$i(t_{k+1}) = \frac{i(t_k)\rho(t_k)}{i'(t_k) + (\rho(t_k) - i'(t_k))e^{-a'\rho(t_k)/(\rho(t_k)-i'(t_k))}}$$

Όπου

$$a' = \frac{\rho(t_k) - i(t_k)}{\rho(t_k)} a(t_k) + \frac{i(t_k)}{\rho(t_k)} a(t_x)$$

Κατά τη σύγκριση των αναλυτικών αποτελεσμάτων με τα αποτελέσματα προσομοιώσεων χρησιμοποιήθηκαν τέσσερα διαφορετικά σενάρια κινητικότητας και δύο ομάδες Bluetooth παραμέτρων.Συνόλικα προέκυψαν οκτώ διαφορετικές περιπτώσεις σύγκρισης κατά τις οποίες τα αναλυτικά αποτελέσματα ταυτίστηκαν πλήρως με τα αποτελέσματα των προσομοιώσεων εκτός από μία περίπτωση που το μοντέλο υπερεκτίμησε την ταχύτητα εξάπλωσης του σκουληκιού κατά την τελευταία φάση εξάπλωσης του.

4.4 Μοντέλα εξάπλωσης XSS σκουληκιών σε online κοινωνικά δίκτυα:

Οι ιστοσελίδες κοινωνικής δικτύωσης ανήκουν στις Web 2.0 υπηρεσίες και είναι ανάμεσα στις πιο επισκέψιμες ιστοσελίδες παγκοσμίως.Η κύρια κατηγορία κοινωνικών δικτύων έχει σαν σκοπό τη σύνδεση φίλων με τις σελίδες ή τα προφίλ που αφορούν τους εαυτούς τους.Συνεπώς οι δημιουργοί κακόβουλου λογισμικού εκμεταλλεύονται την εμπιστοσύνη που υπάρχει ανάμεσα στους χρήστες του κοινωνικού δικτύου για την εξάπλωση αυτού[35].

Το πρώτο ενεργό σκουλήκι που "χτύπησε" τα online κοινωνικά δίκτυα ήταν το MySpace Samy worm το 2005.Το Samy εκμεταλλευόμενο ένα τρωτό σημείο στην ασφάλεια του MySpace ,που ουσιαστικά επρόκειτο για μία cross-site scripting ευπάθεια,ξεκίνησε μολύνοντας ένα άτομο και μετά από είκοσι ώρες κατάφερε να μολύνει 10^6 προφίλ χρηστών.

Το cross-site scripting που αναφέρθηκε προηγουμένως (ή αλλιώς XSS) είναι ένα τρωτό σημείο στην ασφάλεια των περισσότερων εφαρμογών ιστού.Ενώ το XSS είναι ένα κοινό σημείο ευπάθειας των εφαρμογών αυτών,η απειλή του γίνεται ιδιαίτερα αξιοπρόσεκτη λόγω του συνδυασμού HTML και AJAX τεχνολογίας.Η AJAX τεχνολογία επιτρέπει στους πλοηγούς(browsers) να διανέμουν HTTP αιτήσεις εκ μέρους του χρήστη και έτσι ο επιτιθέμενος δεν χρειάζεται να παραπλανήσει το θύμα ώστε να κλικάρει πάνω σε έναν ειδικά φτιαγμένο σύνδεσμο,για παράδειγμα,για να μολυνθεί[36].

Παρακάτω θα αναφερθούν τα βασικά χαρακτηριστικά των XSS σκουληκιών και των κοινωνικών δικτύων καθώς επίσης θα παρουσιαστεί και ένα μοντέλο εξάπλωσης αυτών μαζί με τα αποτελέσματα προσομοίωσης που το επαληθεύουν.

4.4.1 XSS (cross-site scripting) σκουλήκια:

Όπως αναφέρθηκε και προηγουμένως,το cross-site scripting είναι το πλέον συνηθισμένο μέσο,σε επίπεδο εφαρμογής ιστού,που χρησιμοποιούν οι δημιουργοί κακόβουλου λογισμικού για να "κρύψουν" τον κακόβουλο κώδικα στις εφαρμογές.

Υπάρχουν δύο είδη XSS επιθέσεων:

α) Η επίμονη,που είναι γνωστή και σαν αποθηκευμένη (stored) επίθεση,κατά την οποία ο κακόβουλος κώδικας παραμένει αποθηκευμένος μόνιμα στους εξυπηρετητές-θύματα σαν HTML κείμενο (π.χ.μέσα σε βάση δεδομένων,μηνύματα που εστάλθησαν σε forums κλπ).Ο επισκέπτης αποκτά πρόσβαση στον κακόβουλο κώδικα και συνεπώς μολύνεται όταν ανατρέξει στις αποθηκευμένες πληροφορίες μέσω του πλοηγού.

β) Η μη επίμονη,που είναι γνωστή και σαν ανακλαστική (reflective) επίθεση και αποτελεί την πιο συνηθισμένη κατηγορία XSS επιθέσεων.Σε αυτή την κατηγορία επίθεσης,ο κακόβουλος κώδικας στέλνεται στον επισκέπτη εκτός εξυπηρετητή.Για παράδειγμα μέσα σε μήνυμα σφάλματος ή σαν αποτελέσματα αναζήτησης ή σαν οποιαδήποτε απάντηση που περιέχει μέρος ή όλα τα δεδομένα που εστάλθησαν στον εξυπηρετητή σαν μέρος ενός αιτήματος.

Έρευνες έδειξαν ότι περίπου το 80% των εφαρμογών ιστού είναι ευπαθείς σε XSS επιθέσεις.Αυτό συμβαίνει γιατί οι επιτιθέμενοι είναι ελεύθεροι να διοχετεύσουν κακόβουλο κώδικα στις εφαρμογές ιστού μέσω των tags στις φόρμες εισαγωγης[35]. Ένα XSS σκουλήκι ,γνωστό και ως cross-site scripting ιός,είναι κακόβουλος κώδικας που εξαπλώνεται αυτόματα,και χωρίς να παρέμβει ο χρήστης,ανάμεσα στους επισκέπτες μίας ιστοσελίδας.Το XSS σκουλήκι εφαρμογών ιστού,αποτελεί ένα είδος αποθηκευμένης XSS επίθεσης.Πιο συγκεκριμένα αυτός ο τύπος σκουληκιών,έχει την ικανότητα να αντιγράφει τον εαυτό του σε άλλες ιστοσελίδες χρησιμοποιώντας την υπάρχουσα XSS ευπάθεια των εφαρμογών ιστού.

Μία κοινότυπη δράση ενός XSS σκουληκιού είναι η μόλυνση των μελών ενός κοινωνικού δικτύου σε δύο στάδια.Αρχικά ο δημιουργός του σκουληκιού προσθέτει στο προφίλ του το κακόβουλο φορτίο (payload).Στη συνέχεια οποιοσδήποτε επισκεφθεί το μολυσμένο προφίλ,μολύνεται και το κακόβουλο φορτίο προστίθεται στο προφίλ του επισκέπτη μετατρέποντας το σε πηγή μόλυνσης.

Κύριος περιορισμός των XSS σκουληκιών είναι ότι πρέπει να τρέχουν στο περιβάλλον του πλοηγού ιστού.Ακόμα και έτσι όμως είναι ικανά προβάλουν spam μηνυματα ή να προκαλούν DDoS στο κατα περίπτωση θύμα.

4.4.2 Χαρακτηριστικά των κοινωνικών δικτύων:

Υποθέτουμε ότι κάθε χρήστης είναι ένας κόμβος και ότι οι χρήστες που γνωρίζονται μεταξύ τους συνδέονται.

Τα πραγματικά κοινωνικά δίκτυα είναι ομαδοποιημένα small-world δίκτυα με κατανομή βαθμού που συνήθως ακολουθεί power-law.

Τα κύρια χαρακτηριστικά ενός κοινωνικού δικτύου είναι τα παρακάτω:

α) Μικρή μέση απόσταση δικτύου περίπου ίση με logn/logd όπου n είναι το πλήθος των ατόμων στο δίκτυο και d ο μέσος βαθμός του ισοδύναμου (με το δίκτυο) γράφου.

β) Τα κοινωνικά δίκτυα συνήθως παρουσιάζουν υψηλή ομαδοποίηση ή τοπική μεταβατικότητα που σημαίνει ότι αν το άτομο Α γνωρίζει το Β και το Γ τότε είναι πολύ πιθανό οι Β και Γ να γνωρίζονται μεταξύ τους.Ο συντελεστής ομαδοποίησης ενός κόμβου ν είναι C(ν) και ο συντελεστής ομαδοποίησης ενός γράφου είναι ο μέσος όρος των συντελεστών ομαδοποίησης όλων των κόμβων του.Στα πραγματικά κοινωνικά δίκτυα ο συντελεστής ομαδοποίησης κυμαίνεται από 0.1 έως 0.7.

γ) Μία προσεγγιστική power-law κατανομή του βαθμού των κόμβων,κατά την οποία η πιθανότητα για έναν κόμβο ν να έχει βαθμό k (δηλαδή η p(k)) είναι ανάλογη του $k^{-\alpha}$ όπου α είναι ο power-law εκθέτης.

4.4.3 Μαθηματική ανάλυση του μοντέλου εξάπλωσης:

Κατά τη διαδικασία εξάπλωσης ενός XSS σκουληκιού οι χρήστες/προφίλ μπορούν να βρεθούν είτε σε ευάλωτη κατάσταση είτε σε μολυσματική κατάσταση. Συνεπώς η μοντελοποίηση θα γίνει χρησιμοποιώντας το απλό επιδημιολογικό μοντέλο SI το οποίο χαρακτηρίζεται από την εξίσωση (1):

$$\frac{dI(t)}{dt} = \beta \frac{S(t)}{N} I(t)$$

όπου I(t) είναι το πλήθος των μολυσματικών χρηστών σε χρόνο t,S(t) το πλήθος των ευάλωτων χρηστών,N το συνόλικο πλήθος χρηστων και β η παράμετρος μόλυνσης ή ο μέσος ρυθμός επισκέψεων προφίλ που κάνει ένας χρήστης.

Η πιθανότητα ένας ευάλωτος χρήστης να επισκευθεί ένα μολυσματικό προφίλ

είναι I(t)/N άρα ο ρυθμός μόλυνσης ενός ευάλωτου χρήστη είναι βI(t)/N και τελικά ο ρυθμός μόλυνσης ολόκληρου του πληθυσμού ευάλωτων χρηστών είναι βS(t)I(t)/N απ'όπου προκύπτει και η εξίσωση (1) περιγραφής του μοντέλου εξάπλωσης.

Λύνοντας την (1) θεωρώντας αρχική λύση $i(0)=i_0$ έχουμε ότι

$$I(t) = \frac{i_0 N}{i_0 + (N - i_0)e^{-\beta t}}$$

Οι ιστοσελίδες κοινωνικής δικτύωσης,όπως το Facebook και το MySpace, βασίζονται στις σχέσεις φιλίας μεταξύ των χρηστών.Το συγκεκριμένο γεγονός όμως δεν λαμβάνεται υπόψη από το παραπάνω επιδημιολογικό μοντέλο.Για παράδειγμα,κάθε χρήστης επισκέπτεται τα προφίλ των φίλων του πολύ πιο συχνά απ'ότι τα προφίλ όσων δεν είναι φίλοι του.Συνεπώς το β που είναι ο μέσος ρυθμός επισκεψιμότητας ενός προφίλ,είναι συνάρτηση της πιθανότητας επίσκεψης ενός φιλικού προφίλ στα κοινωνικά δίκτυα.

Θεωρούμε q την πιθανότητα ένας χρήστης να επισκεφθεί ένα φιλικό προφίλ.Είναι προφανές ότι όσο αυξάνεται η πιθανότητα q τόσο μεγαλύτερος είναι ο περιορισμός της μόλυνσης.Αυτό συμβαίνει γιατί όταν οι χρήστες επισκέπτονται περισσότερο τους φίλους τους απ'ότι άλλους,ο μολυσματικός πληθυσμός περιορίζεται μεταξύ των φίλων και η μόλυνση θα φτάσει στα υπόλοιπα σημεία του δικτύου με μεγαλύτερη καθυστέρηση.

Μετά την μόλυνση των περισσότερων χρηστών,ο ρυθμός μόλυνσης θα μειωθεί αφού ο συνολικός αριθμός των ευάλωτων χρηστών θα έχει μειωθεί κατά πολύ.Επιπλέον στα κοινωνικά δίκτυα λόγω των σχέσεων φιλίας μεταξύ των μελών και του γεγονότος ότι κάθε χρήστης επισκέπτεται τους φίλους του συχνότερα απ'τους υπόλοιπους,ο συνολικός αριθμός των χρηστών που πιθανόν να επισκεφθεί ένας χρήστης είναι μικρότερος από τον συνολικό αριθμό των χρηστών στο δίκτυο.Άρα η αύξηση της πιθανότητας q στους χρήστες των οποίων οι φίλοι είναι σε μικρότερη αναλογία μολυσμένοι,οδηγεί στην καθυστέρηση της εξάπλωσης του σκουληκιού στο κοινωνικό δίκτυο.Συνεπώς και οι ευάλωτοι χρήστες εξαρτώνται από την πιθανότητα q.

Βάσει των προηγούμενων προτείνεται η εξίσωση (2) της οποίας η ακρίβεια επαληθεύεται μέσω των αποτελεσμάτων της προσομοίωσης που ακολουθεί.

$$\frac{dI(t)}{dt} = \beta(q)\frac{[S(t)]^{(K(q))}}{N} I(t)$$

4.4.4 Προσομοίωση- συμπεράσματα:

Πρωτού περιγραφεί η προσομοίωση θα χρησιμοποιηθεί ένας αλγόριθμος δημιουργίας ενός κοινωνικού δικτύου με τα χαρακτηριστικά που αναφέρθηκαν στην προηγούμενη υποενότητα.Πιο συγκεκριμένα,θα χρησιμοποιηθεί ο αλγόριθμος των Holme και Beom [37].

Δημιουργία κοινωνικού δικτύου:

Χρησιμοποιώντας τον αλγόριθμο που αναφέρθηκε προηγουμένως θα δημιουργηθεί ένας γράφος με τα χαρακτηριστικά των κοινωνικών δικτύων με power-law εκθέτη α=3.

Οι παράμετροι που θα χρησιμοποιηθούν είναι n=10000,m=m_0=3 και m_t=1.8.Οι παράμετροι του παραγόμενου γράφου παρουσιάζονται στον παρακάτω πίνακα

Graph Parameter	Value
Number of Vertices	10000
Number of Edges	29990
Clustering Coefficent	0.1409392
Average Shortest Path	5.133096
Maximum Degree	190
Longest Path (Diameter)	10
Degrees Average (d)	5.998
$\frac{\log n}{\log d}$	5.1413

Για τις ανάγκες της προσομοίωσης των επιρροών των τοπολογικών παραμέτρων στην εξάπλωση του σκουληκιου (όπως ο συντελεστής ομαδοποίησης), είναι απαραίτητη η δημιουργία ενός τυχαίου γράφου παρόμοιου με το παραπάνω κοινωνικό δίκτυο.Αυτό θα επιτευχθεί με τη χρήση του αλγορίθμου των Viger και Latapy [38].

Ο παραγόμενος τυχαίος γράφος έχει την ίδια κατανομή βαθμού με το κοινωνικό δίκτυο που δημιουργήθηκε αρχικά αλλά διαφορετικό συντελεστή ομαδοποίησης.Στον παρακάτω πίνακα παρουσιάζονται οι παράμετροι του παραγόμενου τυχαίου γράφου.

Graph Parameter	Value
Number of Vertices	10000
Number of Edges	29990
Clustering Coefficent	0.003581474
Average Shortest Path	4.407071
Maximum Degree	190
Longest Path (Diameter)	8
Degrees Average (d)	5.998
$\frac{\log n}{\log d}$	5.1413

Όπως φαίνεται στον παραπάνω πίνακα ο συντελεστής ομαδοποίησης του κοινωνικού δικτύου είναι 40 φορές μεγαλύτερος από αυτόν του τυχαίου γράφου,γεγονός που αντικατοπτρίζει την υψηλή ομαδοποίηση των κοινωνικών δικτύων.

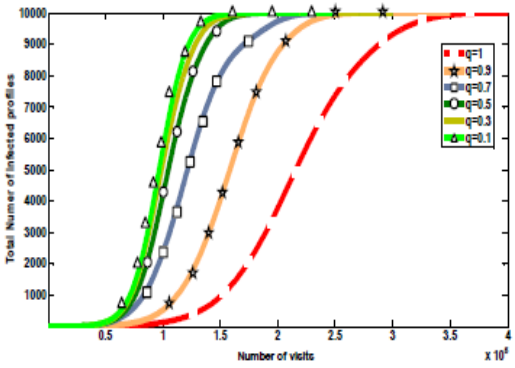
Περιγραφή μοντέλου προσομοίωσης:

Κατά την διαδικασία εξάπλωσης ενός XSS σκουληκιού,ένας ευάλωτος χρήστης πρέπει να επισκεφθεί ένα μολυσματικό προφίλ έτσι ώστε να μολυνθεί.Η ευπάθεια του χρήστη καθορίζεται από το αν ο πλοηγός ιστού του χρήστη μπορεί να εκτελέσει το κακόβουλο script ή όχι.Η πιθανή αδυναμία να εκτελεστεί το script οφείλεται σε ενέργειες που μπορεί να έχουν κάνει οι χρήστες όπως για παράδειγμα η εγκατάσταση συγκεκριμένων πρόσθετων (add-ons) προγραμμάτων στον πλοηγό που αποτρέπουν την εκτέλεση script.

Όπως αναφέρθηκε προηγουμένως,αν ένας ευάλωτος χρήστης επισκεφθεί ένα μολυσματικό προφίλ τότε μολύνεται.Το άτομο,του οποίου το προφίλ δέχεται την επίσκεψη,μπορεί να είναι είτε φίλος του χρήστη-επισκέπτη είτε να μην είναι ένας από αυτούς.Συνεπώς με q_i ορίζεται η πιθανότητα ένας χρήστης i να επισκεφθεί κάποιο φιλικό προφίλ.

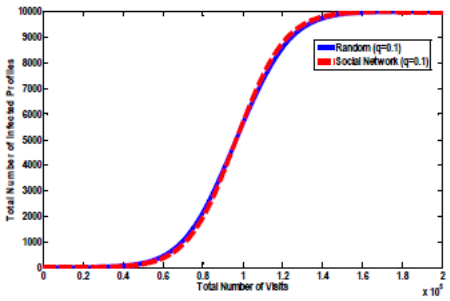
Πριν προχωρήσουμε στην προσομοίωση αξίζει να σημειωθεί ότι όλοι οι χρήστες του δικτύου θεωρούνται ευάλωτοι.

Κάθε άτομο επιλέγεται με ίση πιθανότητα από το σύνολο χρηστών του δικτύου.Το επιλεγμένο άτομο επισκέπτεται ένα προφίλ που ανήκει σε φίλο του με πιθανότητα q και ένα οποιοδήποτε άλλο προφίλ με πιθανότητα 1-q.Η πιθανότητα q είναι ίση για όλα τα άτομα του δικτύου[35].

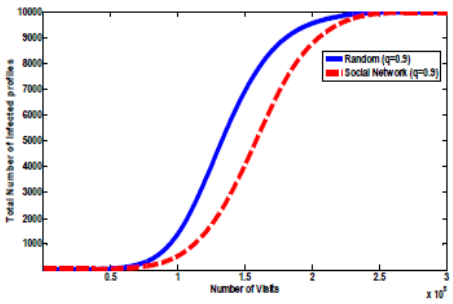


Σχ.13 Εξάπλωση για διαφορετικό q

Όπως είναι εμφανές στην γραφική παράσταση του Σχ.13, όσο συχνότερα κάθε άτομο επισκέπτεται φιλικά προφίλ,τόσο η εξάπλωση του σκουληκιού καθυστερεί.Στα Σχ.14 και Σχ.15 προσομοιώνεται η εξάπλωση ενός XSS σκουληκιού σε ένα κοινωνικό και ένα ισοδύναμο τυχαίο δίκτυο με πιθανότητες q=0.1 και q=0.9.Με τον τρόπο αυτό παρουσιάζεται η επιρροή του συντελεστή ομαδοποίησης στην εξάπλωση του σκουληκιού στο κοινωνικό δίκτυο[35].



Σχ.14 Τυχαίο VS Κοινωνικό Δίκτυο για q=0.1



Σχ.15 Τυχαίο VS Κοινωνικό Δίκτυο για q=0.9

Παρόλο που και τα δύο δίκτυα έχουν την ίδια κατανομή βαθμού,αντικατοπτρίζουν διαφορετική εξάπλωση του σκουληκιού για μεγαλύτερη πιθανότητα q.Ένας λόγος που συμβαίνει αυτό είναι ότι η τοπολογία του δικτύου δεν παίζει σημαντικό ρόλο όταν η πιθανότητα επίσκεψης φίλων είναι μικρή.Χαμηλή τέτοια πιθανότητα σημαίνει ότι οι χρήστες επιλέγουν τυχαία να επισκεφθούν κάποιο προφίλ,με αποτέλεσμα η εξάπλωση του σκουληκιού να είναι γρηγορότερη.

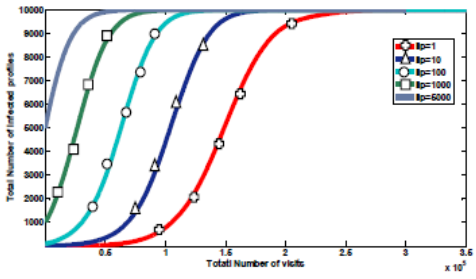
Μία από τις βασικές διαφορές ανάμεσα στο κοινωνικό και το τυχαίο δίκτυο

είναι ότι το κοινωνικό είναι υψηλα ομαδοποιημένο με αποτέλεσμα η εξάπλωση του σκουληκιού να καθυστερείται για μεγάλες τιμές του q.Η καθυστέρηση οφείλεται στο παρακάτω γεγονός.

Σε ένα υψηλά ομαδοποιημένο δίκτυο με μεγάλο q μπορούμε να υποθέσουμε δύο περιπτώσεις:

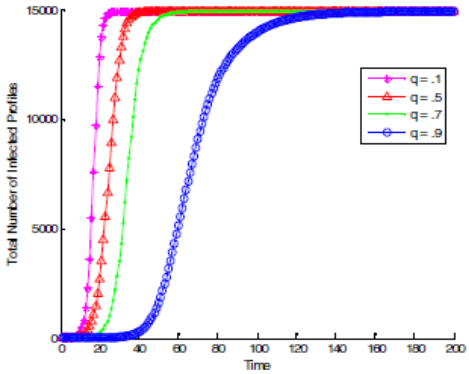
α) ότι η ομάδα (cluster) έχει ένα τουλάχιστον μολυσματικό μέλος.Σε αυτή την περίπτωση,τα υπόλοιπα μέλη της ομάδας θα μολυνθούν σε μικρό χρονικό διάστημα,αλλά η μόλυνση θα περιοριστεί εντός της ομάδας και συνεπώς θα καθυστερήσει η εξάπλωση στα υπόλοιπα μέρη του δικτύου.

β) ότι η ομάδα δεν έχει κανένα μολυσματικό μέλος.Σε αυτή την περίπτωση,αν κάποιο μέλος της ομάδας επισκεφθεί κάποιο προφίλ,με μεγάλη πιθανότητα q θα επιλέξει κάποιο από τα υπόλοιπα μέλη της ομάδας και συνεπώς δεν θα υπάρξει μόλυνση.Άρα το συγκεκριμένο είδος τοπολογίας καθυστερεί σημαντικά την εξάπλωση του σκουληκιού.



Σχ.16 Διαφορετικές τιμές iip για q=0.9

Στο Σχ.16 παρουσιάζεται η επιρροή των αρχικά μολυσμένων προφίλ (iip) στην εξάπλωση ενός XSS σκουληκιού.Στο Σχ.4 επιλέχθηκαν διαφορετικές τιμές για το πλήθος των αρχικά μολυσματικών προφίλ δηλαδή iip= 1,10, 100,1000,5000 με q=0.9 και όπως φαίνεται από τις γραφικές παραστάσεις, η αύξηση τους προκαλεί δραματική αύξηση στην ταχύτητα εξάπλωσης του σκουληκιού.Τέλος για την ανάγκη επαλήθευσης της ακρίβειας της εξίσωσης (2) θα δημιουργηθεί ένα κοινωνικό δίκτυο με βάση την μελέτη του Thelwall[39] με n=15000 πλήθος χρηστών. Στο Σχ.17 παρουσιάζεται η εξάπλωση ενός XSS σκουληκιού για τέσσερις διαφορετικές τιμές τιμές της πιθανότητας q.

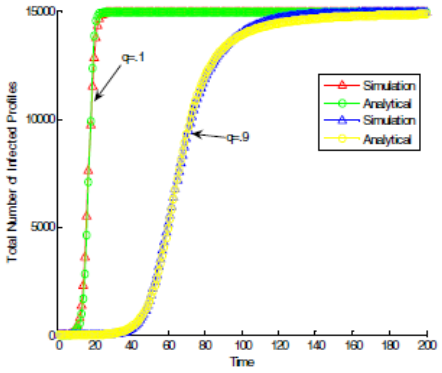


Σχ.17 Αποτελέσματα προσομοίωσης εξάπλωσης XSS σκουληκιού

Κάνοντας interpolation των γραφικών παραστάσεων του Σχ.17 ,τα αποτελέσματα της προσομοίωσης μπορούν να προσαρμοστούν χρησιμοποιώντας τις παρακάτω εξισώσεις

$$\beta(q) = \frac{0.03989}{q^3 - 1.078q^2 + 0.4052q + 0.02036}$$
$$K(q) = 1 + 0.55q$$

και έτσι προκύπτει η γραφική παράσταση των αποτελεσμάτων του προτεινόμενου μοντέλου (Σχ.18 Analytical) για q=0.1 και q=0.9 και παρουσιάζεται η ταύτιση αυτών με τα αποτελεσμάτα της προσομοίωσης.Στο Σχ. 18 αποδεικνύεται η ακρίβεια της εξίσωσης (2)[36].



Σχ.18 Σύγκριση μοντέλου με προσομοίωση

4.4.5. Σύνοψη:

- Faghani, Saidi (2009 [35])
- Ένα XSS σκουλήκι ,γνωστό και ως cross-site scripting ιός,είναι κακόβουλος κώδικας που εξαπλώνεται αυτόματα,και χωρίς να παρέμβει ο χρήστης,ανάμεσα στους επισκέπτες μίας ιστοσελίδας.
- Μία κοινότυπη δράση ενός XSS σκουληκιού είναι η μόλυνση των μελών ενός κοινωνικού δικτύου σε δύο στάδια.

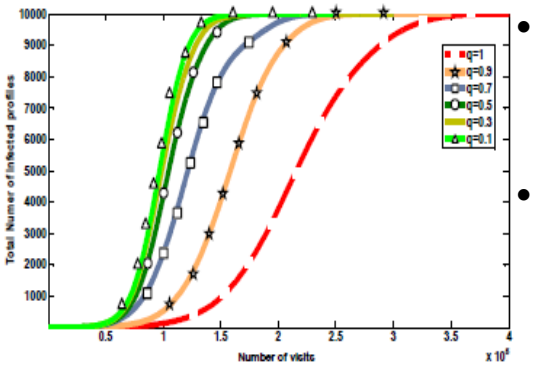
  1)ο δημιουργός του σκουληκιού προσθέτει στο προφίλ του το κακόβουλο φορτίο (payload).

2)οποιοσδήποτε επισκεφθεί το μολυσμένο προφίλ,μολύνεται και το κακόβουλο φορτίο προστίθεται στο προφίλ του επισκέπτη μετατρέποντας το σε πηγή μόλυνσης.

- Για την μοντελοποίηση χρησιμοποιείται τροποποιημένο SI μοντέλο(1).

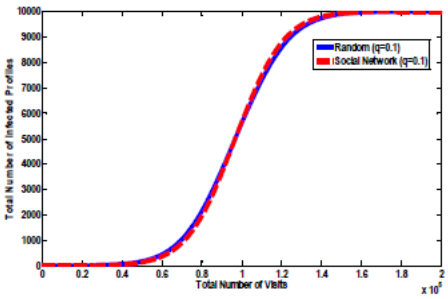
$$\frac{dI(t)}{dt} = \beta(q)\frac{[S(t)]^{(K(q))}}{N} I(t)$$

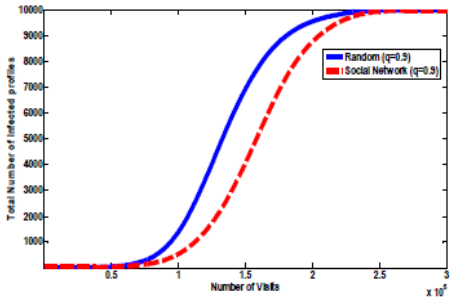
- β είναι ο μέσος ρυθμός επισκεψιμότητας ενός προφίλ και συνάρτηση της πιθανότητας επίσκεψης ενός φιλικού προφίλ q στα κοινωνικά δίκτυα.



- όσο αυξάνεται η πιθανότητα q τόσο μεγαλύτερος είναι ο περιορισμός της μόλυνσης.
- η αύξηση της πιθανότητας q στους χρήστες των οποίων οι φίλοι είναι σε μικρότερη αναλογία μολυσμένοι, οδηγεί στην καθυστέρηση της εξάπλωσης του σκουληκιού στο κοινωνικό δίκτυο.Συνεπώς και οι ευάλωτοι χρήστες εξαρτώνται από την πιθανότητα q.

- Όσο συχνότερα κάθε άτομο επισκέπτεται φιλικά προφίλ(μεγαλύτερο q),τόσο η εξάπλωση του σκουληκιού καθυστερεί. Προσομοίωση εξάπλωσης ενός XSS





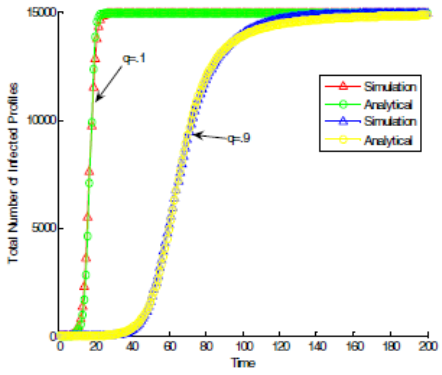
σκουληκιού σε ένα κοινωνικό και ένα ισοδύναμο τυχαίο δίκτυο με πιθανότητες q=0.1 και q=0.9

Τα δύο δίκτυα έχουν την ίδια κατανομή βαθμού,όμως αντικατοπτρίζουν διαφορετική εξάπλωση του σκουληκιού για μεγαλύτερη πιθανότητα q.Αυτό συμβαίνει γιατί η τοπολογία του δικτύου δεν παίζει σημαντικό ρόλο όταν η

πιθανότητα επίσκεψης φίλων είναι μικρή.Χαμηλή τέτοια πιθανότητα σημαίνει ότι οι χρήστες επιλέγουν τυχαία να επισκεφθούν κάποιο προφίλ,με αποτέλεσμα η εξάπλωση του σκουληκιού να είναι γρηγορότερη.

Μία από τις βασικές διαφορές ανάμεσα στο κοινωνικό και το τυχαίο δίκτυο είναι ότι το κοινωνικό είναι υψηλα ομαδοποιημένο με αποτέλεσμα η εξάπλωση του σκουληκιού να καθυστερείται για μεγάλες τιμές του q

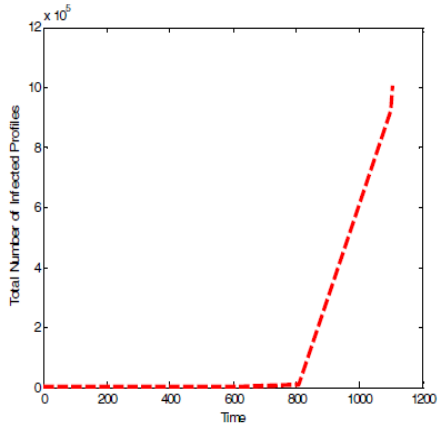
- Τέλος, κατά τη σύγκριση αναλυτικών αποτελεσμάτων και αποτελεσμάτων προσομοίωσης για q=0.1 και q=0.9 αποδεικνύεται η μεγάλη ακρίβεια της εξίσωσης (1).



5.Προτεινόμενο μοντέλο εξάπλωσης για XSS σκουλήκια σε online κοινωνικά δίκτυα.

5.1 Πρόταση

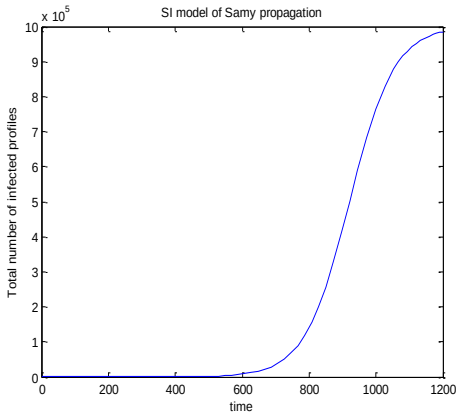
Στο συγκεκριμένο κεφάλαιο της διπλωματικής εργασίας θα χρησιμοποιηθεί το μοντέλο εξάπλωσης two-factor,κατάλληλα τροποποιημένο,ώστε να περιγραφεί η διαδικασία εξάπλωσης των XSS σκουληκιών σε online κοινωνικά δίκτυα.



Σχ.19 Εξάπλωση σκουληκιού Samy

Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο,το πρώτο ενεργό σκουλήκι που έκανε την εμφάνισή του σε online κοινωνικό δίκτυο ήταν το Samy worm στο δίκτυο Myspace το 2005.Το συγκεκριμένο σκουλήκι χρειάστηκε μόλις 20 ώρες για την μόλυνση 10^6 προφίλ χρηστών,ξεκινώντας από ένα και μοναδικό μολυσματικό προφίλ.

Η ταχύτατη εξάπλωση του συγκεκριμένου σκουληκιού,είχε σαν αποτέλεσμα την αναστολή λειτουργίας της ιστοσελίδας,έπειτα από δύο μέρες δράσης του σκουληκιού,με σκοπό την επιδιόρθωση του τρωτού σημείου που εκμεταλλευόταν το σκουλήκι.Στο Σχ.19 παρουσιάζεται γραφικά η εκθετική εξάπλωση του σκουληκιού βάσει πληροφοριών του [40].



Σχ.20 Εξάπλωση σκουληκιού Samy βάσει μοντέλου SI

Για την μοντελοποίηση του σκουληκιού Samy έχει ήδη χρησιμοποιηθεί το απλό επιδημικό μοντέλο SI σύμφωνα με το οποίο η εξάπλωση του συγκεκριμένου σκουληκιού βάσει προσομοίωσης στο Matlab φαίνεται στο παρακάτω σχήμα(Σχ.20).

Κατά την παραπάνω προσομοίωση χρησιμοποιήθηκε η εξίσωση 4.4.3.(1) με μέγεθος πληθυσμού $N=10^6$ που θεωρείται στο σύνολο του ευάλωτος,έναν αρχικά μολυσματικό κόμβο και ρυθμό μόλυνσης β=0.015/N.Παρόλ'αυτά πρέπει να ληφθούν υπόψη μερικά πολύ σημαντικά στοιχεία τα οποία καθιστούν το μοντέλο SI μη ρεαλιστικό για την μοντελοποίηση της διαδικασίας εξάπλωσης ενός αντίστοιχου σκουληκιού.

α) Ο ρυθμός μόλυνσης β δεν είναι σταθερός,αλλά εξαρτάται από την πιθανότητα επίσκεψης φιλικού προφίλ q και στην ουσία έχει πτωτική τάση όσο το q αυξάνεται.Το συγκεκριμένο γεγονός έχει ήδη ληφθεί υπόψη από τους Faghani και Saidi [36] όπως έχουμε ήδη προαναφέρει,οι οποίοι προσάρμοσαν την συγκεκριμένη πιθανότητα q στο επιδημικό μοντέλο SI και μοντελοποίησαν την εξάπλωση του σκουληκιού χρησιμοποιώντας κάθε φορά τυχαία τιμή της πιθανότητας. Ο προσδιορισμός της πιθανότητας μπορεί να γίνει βάσει καθορισμένου τύπου ώστε να βελτιωθεί η ακρίβεια του μοντέλου,αφού είναι προφανές ότι η πιθανότητα εξαρτάται από τον αριθμό των συνδέσεων ενός κόμβου δηλαδή από τον βαθμό του κόμβου.

Αυτό συμβαίνει γιατί αν πάρουμε για παράδειγμα ένα online κοινωνικό δίκτυο σαν το MySpace το όποιο δεν είναι ομογενές(όπως και όλα τα κοινωνικά δίκτυα),όσο περισσότερα είναι τα φιλικά προφίλ ενός χρήστη i(δηλαδή όσο μεγαλύτερο το k_i),τόσο μεγαλύτερη η πιθανότητα ο χρήστης αυτός να επισκεφθεί κάποιο από αυτά παρά κάποιο άλλο μη φιλικό(μεγαλύτερο q_i).Συνεπώς η πιθανότητα q δεν είναι σταθερή για όλους τους χρήστες και αυτό βασίζεται στην ανομοιογενή φύση των κοινωνικών δικτύων.

Ένας τρόπος για τον προσδιορισμό της πιθανότητας q προκύπτει από το

ακόλουθο παράδειγμα.Έστω ένας χρήστης i ενός online κοινωνικού δικτύου,ο οποίος συνδέεται με k_i άλλους χρήστες από το σύνολο N όλων των χρηστών του δικτύου.Τότε η πιθανότητα q_i,δηλαδή η πιθανότητα ο χρήστης αυτός να επισκεφθεί οποιονδήποτε από αυτούς τους φίλους,δίνεται από τον τύπο $q_i=k_i/N$.Συνεπώς μία προσεγγιστική ενιαία τιμή για την πιθανότητα q θα μπορούσε να προκύψει από τον τύπο q=<k>/N όπου <k> είναι ο μέσος βαθμός του δικτύου.

β) Στο SI μοντέλο οι κόμβοι μπορούν να μεταβούν μόνο από την ευάλωτη κατάσταση στην μολυσματική κατάσταση.Με βάση λοιπόν τη θεώρηση ότι η εξάπλωση ενός XSS σκουληκιού περιγράφεται κατάλληλα από το συγκεκριμένο μοντέλο, παραλείπονται πλήρως οι συνέπειες των ανθρώπινων αντιμέτρων στην διαδικασία εξάπλωσης.Συνεπώς μία πιο ρεαλιστική προσέγγιση μπορεί να επιτευχθεί με τη χρήση του μοντέλου two-factor με τη διαφορά ότι πρέπει να χρησιμοποιηθεί και ο πρώτος συλλογισμός (ότι δηλαδή ο ρυθμός μόλυνσης είναι συνάρτηση της πιθανότητας q).

Η καταλληλότητα του μοντέλου two factor για τον συγκεκριμένο σκοπό αποδεικνύεται από το γεγονός ότι όπως και σε όλα τα είδη σκουληκιών έτσι και στα XSS σκουλήκια,τα ανθρώπινα αντίμετρα επιρρεάζουν σημαντικά την εξάπλωση τους.

Στην περίπτωση των XSS σκουληκιών,η εγκατάσταση add-on προγραμμάτων στον πλοηγό ιστού που χρησιμοποιεί ο χρήστης,με σκοπό την αποτροπή εκτέλεσης script συνεπάγεται ότι ο χρήστης είναι αδύνατο να μολυνθεί,μιάς και ο πλοηγός είναι το μέσο με το οποίο ο κακόβουλος κώδικας μεταδίδεται σε αυτόν.Επιπλέον η εγκατάσταση ή η ενημέρωση ενός αντιικού λογισμικού συμβάλλει στην ανίχνευση του κακόβουλου κώδικα και πολλές φορές στην ανοσοποίηση του χρήστη πριν ακόμα αυτός μολυνθεί.

Όλα τα παραπάνω,είναι προφανές ότι καθυστερούν την εξάπλωση του σκουληκιού και πρέπει οπωσδήποτε να ληφθούν υπόψη κατά την μοντελοποίηση της. Το μοντέλο two factor όπως είδαμε και στο κεφάλαιο 3.8 περιγράφεται από το σύστημα διαφορικών εξισώσεων

$$\frac{dS(t)}{dt} = -\beta(t)S(t)I(t) - \frac{dQ(t)}{dt}$$
$$\frac{dR(t)}{dt} = \gamma I(t)$$
$$\frac{dQ(t)}{dt} = \mu S(t)J(t)$$
$$\beta(t) = \beta_0[1 - \frac{I(t)}{N}]^n$$
$$N = S(t) + I(t) + R(t) + Q(t)$$
$$I(0) = I_0 << N \ , \ S(0) = N - I_0 \ , \ R(0) = Q(0) = 0$$

Το μόνο που πρέπει να τροποποιηθεί είναι ο τύπος του ρυθμού μόλυνσης β(t),όπου πρέπει να συμπεριληφθεί η πιθανότητα q.

Μία σκέψη είναι να πολλαπλασιαστεί η πιθανότητα q με τον εκθέτη n,ο οποίος χρησιμοποιείται για την προσαρμογή της ευαισθησίας του ρυθμού μόλυνσης στον αριθμό των μολυσματικών κόμβων I(t).

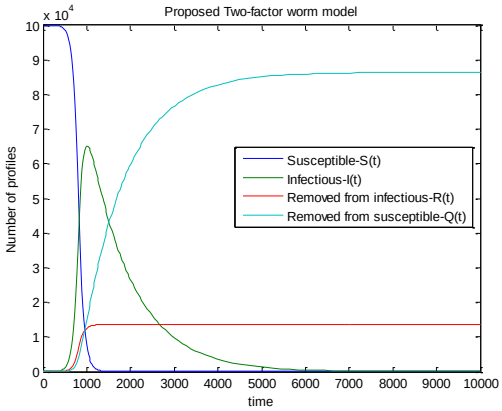
Σαν αποτέλεσμα προκύπτει ο τύπος $\beta(t) = \beta_0[1 - \frac{I(t)}{N}]^{n*q}$ σύμφωνα με τον οποίο όταν αυξάνεται η πιθανότητα q,ο ρυθμός μόλυνσης β(t) μειώνεται.

5.2 Περιγραφή προσομοίωσης-Συμπεράσματα

Αρχικά χρησιμοποιούνται στοιχεία από το [41] που παρέχονται πληροφορίες για τα τοπολογικά χαρακτηριστικά του κοινωνικού δικτύου MySpace.Συνεπώς στις παρακάτω προσομοιώσεις, θεωρούμε δείγμα πληθυσμού 100000 χρηστών-προφιλ(κόμβων) και μέσο βαθμό δικτύου <k>=137.1 .Βάσει του τύπου υπολογισμού της πιθανότητας q που παρουσιάστηκε στην προηγούμενη υποενότητα έχουμε ότι $q=<k>/N=137.1/100000=1.371*10^{-3}$.

Χρησιμοποιώντας το τροποποιημένο two factor μοντέλο που προτάθηκε και ρυθμούς μόλυνσης,απομάκρυνσης ευάλωτων κόμβων και απομάκρυνσης μολυσμένων κόμβων, β=0.015/N, μ=0.002/N και γ=0.001 αντίστοιχα προκύπτουν οι παραπάνω

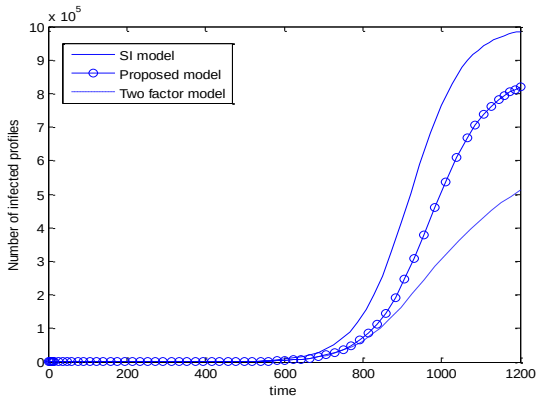
γραφικές παραστάσεις (Σχ.21) με αρχικό πλήθος κόμβων I(0)=1,S(0)=N-I(0) και R(0)=Q(0)=0.



Σχ.21 Τροποποιημένο Two factor μοντέλο

Χρησιμοποιώντας τις ίδιες ακριβώς παραμέτρους για τα τρία μοντέλα προκύπτει το παρακάτω σχήμα (Σχ.22) στο οποίο φαίνεται η αύξηση του αριθμού των μολυσμένων κόμβων(J(t)=R(t)+I(t) στην περίπτωση του two factor και του προτεινόμενου μοντέλου) τις πρώτες 20 ώρες από την απελευθέρωση του σκουληκιού.Οπως είναι εμφανές,το προτεινόμενο μοντέλο προσεγγίζει την διαδικασία εξάπλωσης του σκουληκιού Samy(και κατ'επέκταση οποιουδήποτε XSS σκουληκιού)με μεγαλύτερη ακρίβεια από το κλασσικό two factor μοντέλο.

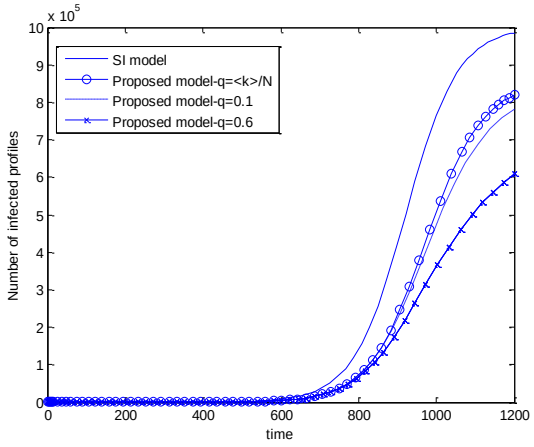
Για να διαπιστώσουμε την σημασία σωστού προσδιορισμού της πιθανότητας επίσκεψης φιλικού προφίλ q τρέχουμε ξανά την παραπάνω προσομοίωση χρησιμοποιώντας στη θέση του απλού two factor μοντέλου,το προτεινόμενο μοντέλο με q=0.1 και q=0.6 (αντιπροσωπευτικές τιμές που χρησιμοποιήθηκαν στο [36]).



Σχ.22 Σύγκριση μοντέλων εξάπλωσης XSS σκουληκιού

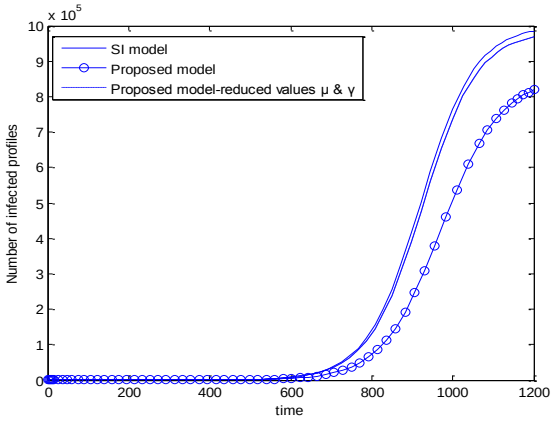
Όπως είναι εμφανές (Σχ.23) και στις δύο αυτές περιπτώσεις η καμπύλη εξάπλωσης του σκουληκιού υπολείπεται σε ακρίβεια απ'ότι αν χρησιμοποιηθεί ο τύπος που ορίστηκε προηγουμένως (q=<k>/N).Τονίζουμε ότι οι τιμές των παραμέτρων είναι ίδιες για όλες τις περιπτώσεις(εκτός του q φυσικά).

Είναι προφανές ότι όσο μειώνεται ο ρυθμός απομάκρυνσης ευάλωτων και μολυσματικών κόμβων τόσο η εξάπλωση που περιγράφεται από το προτεινόμενο



Σχ.23 Σύγκριση προτεινόμενου μοντέλου για διαφορετικές τιμές q

μοντέλο προσεγγίζει την καμπύλη εξάπλωσης βάσει του SI μοντέλου.Αυτό είναι λογικό καθώς με την μείωση των συγκεκριμένων παραμέτρων(μ και γ αντίστοιχα), μειώνεται και το πλήθος των απομακρυσμένων ευάλωτων και μολυσμένων κόμβων(Q και R αντίστοιχα).Το παραπάνω παριστάνεται γραφικά στο Σχ.24 όπου για την δημιουργία της γραφικής παράστασης (‘--‘) οι τιμές του μ και του γ μειώθηκαν και έγιναν μ=0.0002/N και γ=0.0001.



Σχ.24 Σύγκριση προτεινόμενου μοντέλου για διαφορετικές τιμές μ&γ

Στην πραγματικότητα το προτεινόμενο μοντέλο είναι πολύ πιο ρεαλιστικό από το μοντέλο SI που έχει ήδη προταθεί και αυτό λόγω του γεγονότος ότι λαμβάνεται υπόψιν η επιρροή των ανθρώπινων αντιμέτρων στην διαδικασία εξάπλωσης του σκουληκιού σε συνδυασμό με τον πτωτικό ρυθμό μόλυνσης,ο οποίος εξαρτάται από την πιθανότητα q επίσκεψης φιλικού προφίλ.

5.3 Σύνοψη:

Προτεινόμενο μοντέλο εξάπλωσης για XSS σκουλήκια σε online κοινωνικά δίκτυα:

- Πρόταση 1:

  Όσο περισσότερα είναι τα φιλικά προφίλ ενός χρήστη i(δηλαδή όσο μεγαλύτερο το k_i),τόσο μεγαλύτερη η πιθανότητα ο χρήστης αυτός να επισκεφθεί κάποιο από αυτά,παρά κάποιο άλλο μη φιλικό(μεγαλύτερο q_i).Συνεπώς η πιθανότητα q δεν είναι σταθερή για όλους τους χρήστες και αυτό βασίζεται στην ανομοιογενή φύση των κοινωνικών δικτύων.

  Προσέγγιση:

  Μία προσεγγιστική ενιαία τιμή για την πιθανότητα q θα μπορούσε να προκύψει από τον τύπο q=<k>/N όπου <k> είναι ο μέσος βαθμός του δικτύου.

- Πρόταση 2:

  Με τη χρήση του SI μοντέλου παραλείπονται πλήρως οι συνέπειες των ανθρώπινων αντιμέτρων στην διαδικασία εξάπλωσης.Συνεπώς μία πιο ρεαλιστική προσέγγιση μπορεί να επιτευχθεί με τη χρήση του μοντέλου two-factor σε συνδυασμό με την Πρόταση 1.

  Προσέγγιση:

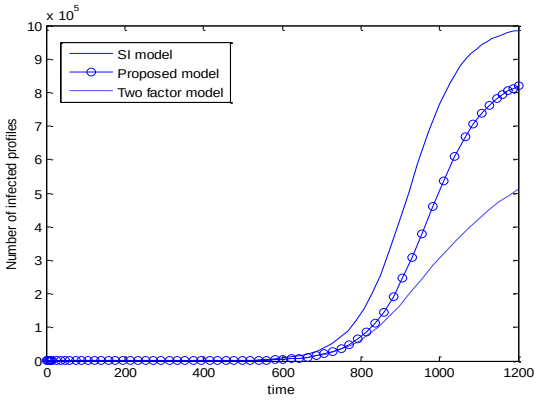
$$\frac{dS(t)}{dt} = -\beta(t)S(t)I(t) - \frac{dQ(t)}{dt}$$
$$\frac{dR(t)}{dt} = \gamma I(t)$$
$$\frac{dQ(t)}{dt} = \mu S(t)J(t)$$
$$\beta(t) = \beta_0[1 - \frac{I(t)}{N}]^n$$
$$N = S(t) + I(t) + R(t) + Q(t)$$
$$I(0) = I_0 << N \ , \ S(0) = N - I_0 \ , \ R(0) = Q(0) = 0$$

  Και σε συνδυασμό με την Πρόταση 1 έχουμε ότι $\beta(t) = \beta_0[1 - \frac{I(t)}{N}]^{n*q}$

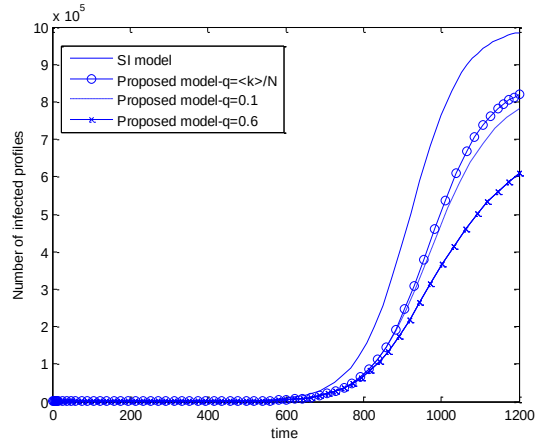
- Χρησιμοποιώντας στοιχεία από το [41] που παρέχονται πληροφορίες για τα τοπολογικά χαρακτηριστικά του κοινωνικού δικτύου MySpace θεωρούμε δείγμα

πληθυσμού 100000 χρηστών-προφιλ(κόμβων) και μέσο βαθμό δικτύου <k>=137.1 και συνεπώς q=<k>/N=137.1/100000=1.371*10^{-3} .

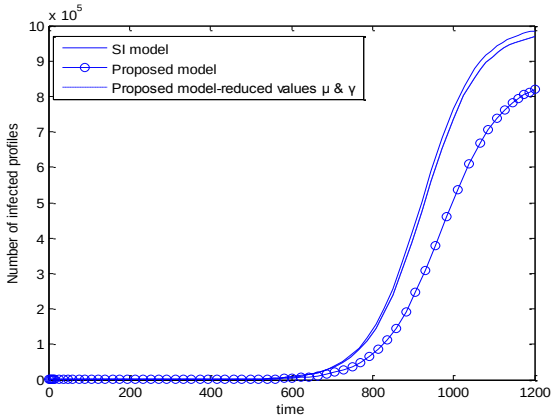
- Σύγκριση απλού μοντέλου SI με το κλασσικό two factor μοντέλο και το προτεινόμενο μοντέλο.Είναι εμφανές ότι το προτεινόμενο μοντέλο προσεγγίζει την διαδικασία εξάπλωσης του σκουληκιού Samy (και κατ'επέκταση οποιουδήποτε XSS σκουληκιού) με μεγαλύτερη ακρίβεια από το κλασσικό two factor μοντέλο.



- Με την παρακάτω προσομοίωση διαπιστώνεται η σημασία σωστού προσδιορισμού της πιθανότητας επίσκεψης φιλικού προφίλ q. Στις δύο περιπτώσεις όπου q=0.1 και q=0.6 η καμπύλη εξάπλωσης του σκουληκιού υπολείπεται σε ακρίβεια απ'ότι αν χρησιμοποιηθεί ο τύπος που ορίστηκε προηγουμένως (q=<k>/N).



- Όσο μειώνεται ο ρυθμός απομάκρυνσης ευάλωτων και μολυσματικών κόμβων τόσο η εξάπλωση που περιγράφεται από το προτεινόμενο μοντέλο προσεγγίζει την καμπύλη εξάπλωσης βάσει του SI μοντέλου.



- Η απόκλιση των αποτελεσμάτων προσομοίωσης του προτεινόμενου μοντέλου (Σχ.22) από αυτά του SI μοντέλου δικαιολογούνται πλήρως από την επιλογή ίδιου μεγέθους πληθυσμού και στις δύο περιπτώσεις.Το προτεινόμενο μοντέλο είναι πολύ πιο ρεαλιστικό από το μοντέλο SI που έχει ήδη προταθεί και αυτό λόγω του γεγονότος ότι λαμβάνεται υπόψιν η επιρροή των ανθρώπινων αντιμέτρων στην διαδικασία εξάπλωσης του σκουληκιού σε συνδυασμό με τον ορθότερο υπολογισμό της πιθανότητας q επίσκεψης φιλικού προφίλ.

Παράρτημα:

Οι γραφικές παραστάσεις που φαίνονται στα Σχημάτα 20,21,22,23 και 24 υλοποιήθηκαν σε Matlab.Οι κώδικες που χρησιμοποιήθηκαν παρατίθενται στο συγκεκριμένο κεφάλαιο.

- Σχήμα 20:

-Αρχείο si.m:

```
clear;
to = 0;
tf =1200;
%tf = 10000;
yo = [999999 1];
[t y] = ode45('ypsi',[to tf],yo);
plot(t,y(:,2))
title('SI model of Samy propagation')
xlabel('time')
ylabel('Total number of infected profiles')
```

που χρησιμοποιεί τη συνάρτηση που ορίζεται στο αρχείο ypsi.m:

```
function ypsi =ypsi(~,y)
a = .015/1000000;

ypsi(1) =-a*y(1)*y(2);
ypsi(2) = a*y(1)*y(2);
ypsi = [ypsi(1) ypsi(2)]';
```

- Σχήμα 21:

-Αρχείο modelprop.m:

```
clear;
to = 0;
%tf = 1200;
tf =10000;
yo = [99999 1 0 0];
[t y] = ode45('ypprop',[to tf],yo);
plot(t,y(:,1),t,y(:,2),t,y(:,3),t,y(:,4))
title('Proposed Two-factor worm model')
xlabel('time')
ylabel('Number of profiles')
legend ('Susceptible-S(t)','Infectious-I(t)','Removed from infec-
tious-R(t)','Removed from susceptible-Q(t)')
```

που χρησιμοποιεί τη συνάρτηση που ορίζεται στο αρχείο ypprop.m:

```
function ypprop =ypprop(~,y)
e = .002/100000;
g = .001;
b0 = 0.015/100000;
q = 1.371*10^(-3); %diko mou skeptiko
%q = 0.6;
ypprop(1) =-(b0*(1-y(2)/100000)^(2*q))*y(1)*y(2)-e*y(1)*(y(2)+y(3));
ypprop(2) =(b0*(1-y(2)/100000)^(2*q))*y(1)*y(2)-g*y(2);
ypprop(3) = e*y(1)*(y(2)+y(3));
ypprop(4) = g*y(2);
ypprop = [ypprop(1) ypprop(2) ypprop(3) ypprop(4)]';
```

- Σχήμα 22:

-Αρχείο sxhma22.m:

```
clear all;
to = 0;
tf =1200;
yo = [999999 1];
[t y] = ode45('ypsi',[to tf],yo);
plot(t,y(:,2)) %SI model
hold on
clear all;
to = 0;
tf =1200;
yo = [999999 1 0 0];
[t y] = ode45('ypprop',[to tf],yo);
plot (t,y(:,2)+y(:,4),'-o') %grafikh twn molusmenwn profil dhl.
J(t)=I(t)+R(t) tou proposed 2factor
hold on
clear all;
to = 0;
tf =1200;
yo = [999999 1 0 0];
[t y] = ode45('yp2fm',[to tf],yo);
plot (t,y(:,2)+y(:,4),'--') %grafikh tou aplou 2 factor
legend ('SI model','Proposed model','Two factor model')
xlabel('time')
ylabel('Number of infected profiles')
```

που χρησιμοποιεί τις συναρτήσεις ypsi,ypprop,yp2fm που ορίζονται στα αρχεία:

-Αρχείο ypsi.m:

```
function ypsi =ypsi(~,y)
a = .015/1000000;

ypsi(1) =-a*y(1)*y(2);
ypsi(2) = a*y(1)*y(2);
ypsi = [ypsi(1) ypsi(2)]';
```

-Αρχείο ypprop.m:

```
function ypprop =ypprop(~,y)
e = .002/1000000;
g = .001;
b0 = 0.015/1000000;
q = 1.37*10^(-4); %diko mou skeptiko
%q = 0.6;
ypprop(1) =-(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-e*y(1)*(y(2)+y(3));
ypprop(2) =(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-g*y(2);
ypprop(3) = e*y(1)*(y(2)+y(3));
ypprop(4) = g*y(2);
ypprop = [ypprop(1) ypprop(2) ypprop(3) ypprop(4)]';
```

-Αρχείο yp2fm.m:

```
function yp2fm =yp2fm(~,y)
e = .002/1000000;
g = .001;
b0 = 0.015/1000000;
%b =@(t) b0*(1-y(2)/1200000)^2;
yp2fm(1) =-(b0*(1-y(2)/1000000)^2)*y(1)*y(2)-e*y(1)*(y(2)+y(3));
yp2fm(2) =(b0*(1-y(2)/1000000)^2)*y(1)*y(2)-g*y(2);
```

```
yp2fm(3) = e*y(1)*(y(2)+y(3));
yp2fm(4) = g*y(2);
yp2fm = [yp2fm(1) yp2fm(2) yp2fm(3) yp2fm(4)]';
```

- Σχήμα 23:

-Αρχείο sxhma23.m:

```
clear all;
to = 0;
tf =1200;
yo = [999999 1];
[t y] = ode45('ypsi',[to tf],yo);
plot(t,y(:,2)) %SI model
hold on
clear all;
to = 0;
tf =1200;
yo = [999999 1 0 0];
[t y] = ode45('ypprop',[to tf],yo);
plot (t,y(:,2)+y(:,4),'-o') %grafikh twn molusmenwn profil dhl.
J(t)=I(t)+R(t) tou proposed 2factor gia q=<k>/N
hold on
clear all;
to = 0;
tf =1200;
yo = [999999 1 0 0];
[t y] = ode45('ypprop1',[to tf],yo);
plot (t,y(:,2)+y(:,4),'--') %grafikh twn molusmenwn profil dhl.
J(t)=I(t)+R(t) tou proposed 2factor gia q=0.1
hold on
clear all;
to = 0;
tf =1200;
yo = [999999 1 0 0];
```

```
[t y] = ode45('ypprop2',[to tf],yo);
plot (t,y(:,2)+y(:,4),'-x') %grafikh twn molusmenwn profil dhl.
J(t)=I(t)+R(t) tou proposed 2factor gia q=0.6
legend ('SI model','Proposed model-q=<k>/N','Proposed model-
q=0.1','Proposed model-q=0.6')
xlabel('time')
ylabel('Number of infected profiles')
```

που χρησιμοποιεί τις συναρτήσεις ypsi,ypprop,ypprop1,ypprop2 που ορίζονται στα αρχεία:

-Αρχείο ypsi.m:

```
function ypsi =ypsi(~,y)
a = .015/1000000;

ypsi(1) =-a*y(1)*y(2);
ypsi(2) = a*y(1)*y(2);
ypsi = [ypsi(1) ypsi(2)]';
```

-Αρχείο ypprop.m:

```
function ypprop =ypprop(~,y)
e = .002/1000000;
g = .001;
b0 = 0.015/1000000;
q = 1.37*10^(-4); %diko mou skeptiko
%q = 0.6;
ypprop(1) =-(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-e*y(1)*(y(2)+y(3));
ypprop(2) =(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-g*y(2);
ypprop(3) = e*y(1)*(y(2)+y(3));
ypprop(4) = g*y(2);
ypprop = [ypprop(1) ypprop(2) ypprop(3) ypprop(4)]';
```

-Αρχείο ypprop1.m:

```
function ypprop1 =ypprop1(~,y)
e = .002/1000000;
g = .001;
b0 = 0.015/1000000;
%q = 1.37*10^(-4); %diko mou skeptiko
q = 0.1;
ypprop1(1) =-(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-
e*y(1)*(y(2)+y(3));
ypprop1(2) =(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-g*y(2);
ypprop1(3) = e*y(1)*(y(2)+y(3));
ypprop1(4) = g*y(2);
ypprop1 = [ypprop1(1) ypprop1(2) ypprop1(3) ypprop1(4)]';
```

-Αρχείο ypprop2.m:

```
function ypprop2 =ypprop2(~,y)
e = .002/1000000;
g = .001;
b0 = 0.015/1000000;
%q = 1.37*10^(-4); %diko mou skeptiko
q = 0.6;
ypprop2(1) =-(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-
e*y(1)*(y(2)+y(3));
ypprop2(2) =(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-g*y(2);
ypprop2(3) = e*y(1)*(y(2)+y(3));
ypprop2(4) = g*y(2);
ypprop2 = [ypprop2(1) ypprop2(2) ypprop2(3) ypprop2(4)]';
```

- Σχήμα 24:

-Αρχείο sxhma24.m:

```
clear all;
to = 0;
tf =1200;
yo = [999999 1];
[t y] = ode45('ypsi',[to tf],yo);
plot(t,y(:,2)) %SI model
hold on
clear all;
to = 0;
tf =1200;
yo = [999999 1 0 0];
[t y] = ode45('ypprop',[to tf],yo);
plot (t,y(:,2)+y(:,4),'-o') %grafikh twn molusmenwn profil dhl.
J(t)=I(t)+R(t) tou proposed 2factor
hold on
clear all;
to = 0;
tf =1200;
yo = [999999 1 0 0];
[t y] = ode45('ypprop1',[to tf],yo);
plot (t,y(:,2)+y(:,4),'--') %grafikh twn molusmenwn profil dhl.
J(t)=I(t)+R(t) tou proposed 2factor gia meiwmenes times m kai g.
legend ('SI model','Proposed model','Proposed model-reduced values μ
& γ')
xlabel('time')
ylabel('Number of infected profiles')
```

που χρησιμοποιεί τις συναρτήσεις ypsi,ypprop,ypprop1 που ορίζονται στα αρχεία:

-Αρχείο ypsi.m:

```
function ypsi =ypsi(t,y)
```

```
a = .015/1000000;

ypsi(1) =-a*y(1)*y(2);
ypsi(2) = a*y(1)*y(2);
ypsi = [ypsi(1) ypsi(2)]';
```

-Αρχείο ypprop.m:

```
function ypprop =ypprop(~,y)
e = .002/1000000;
g = .001;
b0 = 0.015/1000000;
q = 1.37*10^(-4); %diko mou skeptiko
%q = 0.6;
ypprop(1) =-(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-e*y(1)*(y(2)+y(3));
ypprop(2) =(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-g*y(2);
ypprop(3) = e*y(1)*(y(2)+y(3));
ypprop(4) = g*y(2);
ypprop = [ypprop(1) ypprop(2) ypprop(3) ypprop(4)]';
```

-Αρχείο ypprop1.m:

```
function ypprop1 =ypprop1(~,y)
e = .0002/1000000;
g = .0001;
b0 = 0.015/1000000;
q = 1.37*10^(-4); %diko mou skeptiko
%q = 0.1;
ypprop1(1) =-(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-
e*y(1)*(y(2)+y(3));
ypprop1(2) =(b0*(1-y(2)/1000000)^(2*q))*y(1)*y(2)-g*y(2);
ypprop1(3) = e*y(1)*(y(2)+y(3));
ypprop1(4) = g*y(2);
ypprop1 = [ypprop1(1) ypprop1(2) ypprop1(3) ypprop1(4)]';
```

Βιβλιογραφία-Αναφορές:

[1] X.F.*Wang*, G.Chen, Complex networks: Small-world, scale-free, and beyond, IEEE Circuits and Systems Magazine (2003)

[2] P. Erdös, A. Rényi, On the evolution of random graphs, Publ.Math. Inst. Hung. Acad. Sci., vol. 5, pp. 17-60, (1959)

[3] D.J.Watts, S.H.Strogatz, Collective Dynamics of 'small world' networks, Nature,Vol 393, pp 440-442 (1998)

[4] M.E.J.Newman and D.J.Watts, Renormalization group analysis of the small-world network model, Phys. Lett. A, vol. 263, pp. 341-346, (1999)

[5] A-L. Barabási, R. Albert, Emergence of scaling in random networks, *Science*, vol. 286, pp. 509-512, (1999).

[6] A-L. Barabási, R. Albert, H. Jeong, Mean-field theory for scalefree random networks, *Physica A*, vol. 272, pp. 173-187, (1999)

[7] R. Albert, A.-L. Barabási,Statistical mechanics of complex networks. Reviews of Modern Physics 74, 47-97 (2002)

[8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. Hwang,Complex networks: Structure and dynamics. *Physics Reports*, Vol. 424, No. 4-5. (February 2006), pp. 175-308.

[9] M. Faloutsos, P. Faloutsos, C. Faloutsos, Comput. Commun. Rev. 29 (1999) 251

[10] A.S.Tanenbaum, Complex Networks 4th edition.Prentice Hall (2003)

[11] R.A.Grimes, Malicious Mobile Code:Virus Protection for Windows.O'Reilly

(2001)
[12] M.Erbschloe, Trojans, Worms and Spyware:A Computer Security Professional's Guide to Malicious Code. Elsevier Butterworth–Heinemann (2005)
[13] P.Li, M.Salour, Xiao Su: A survey of internet worm detection and containment. IEEE Communications Surveys and Tutorials 10(1-4): 20-35 (2008)
[14] Morris (Computer Worm), retrieved April 2010, http://en.wikipedia.org/wiki/Morris_worm
[15] V. P. D. Moore et al., Inside the Slammer Worm, IEEE Sec.& Privacy, vol.1,(2003) pp. 33–39.
[16] Z.Nikoloski, N.Deo, L.Kucera, Correlation Model of Worm Propagation on Scale-Free Networks, Complexus Network Modelling, 3, 169-182, (2006)
[17] S.Staniford, V.Paxson, V.Weaver: How to own the Internet in your spare time, in Proceedings of the USENIX Security Symposium, Monterey (2002) pp 149–167
[18] N.Weaver, Potential strategies for high-speed active worms: a worst case analysis, (2002)
[19] A.Wagner, T.Dubendorfer, B.Plattner, R.Hiestand:Experiences with worm propagation simulations; in Proceedings of ACM Workshop on Rapid Malcode, Washington, (2003), pp 34–41.
[20] A.Solomon, Epidemiology and computer viruses, http://vx.netlux.org/lib/static/vdat/epepidem.htm (1990)
[21] JO.Kephart,SR.White: Directed-graph epidemiological models of computer viruses, in Proceedings of IEEE Symposium on Security and Privacy, Oakland, (1991),p 343
[22] Y.Wang, C.Wang: Modelling the effects of timing parameters on virus propagation; in Proceedings of ACM Workshop on Rapid Malcode, Washington, (2003) pp 61–66.
[23] R.Pastor-Satorras, A.Vespignani: Epidemics and immunization in scale-free networks; in S.Bornholdt, HG.Schuster (ed): Handbook of Graphs and Networks: from the Genome to the Internet. Weinheim,Wiley-VCH, (2002), pp 113–132.
[24] C.C.Zou, D.Towsley, W.Gong, Code Red Worm Propagation Modeling and Analysis, Proceedings of the 9th ACM Symposium on Computer and

Communications Security, pp. 138-147, (2002)
[25] M.Boguna,R.Pastor-Satorras, A.Vespignani:Epidemic spreading in complex networks with degree correlations; in R.Pastor-Satorras(ed): Lecture Notes in Physics. Berlin , Springer (2003) vol 625, pp 127–147.
[26] MM.Williamson, J.Leveille, An epidemiological model of virus spreading and cleanup, in Proceedings of Virus Bulletin Conference, Montreal , (2003)
[27] Z. Chen, L. Gao, and K. Kwiat, Modeling the Spread of Active Worms, INFOCOM 2003, Twenty-second Annual Joint Conference of the IEEE Computer and Communi-cations Societies, IEEE, vol. 3, pp. 1890-1900, (2003)
[28] G.Serazzi,S.Zanero, Computer virus propagation models, in Tutorials of the 11th IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunications Systems (2003)
[29] A. Kanaoka, E. Okamoto, Propagation Model for a Mass-Mailing Worm with Mailing List, Proceedings of World Academy of Science, Engineering and Technology (WASET), Vol. 36, pp.96-102, (2008)
[30] H.Ebel, L.Mielsch, S.Bornholdt, Scale-free topology of e-mail networks, Physical Review E 66, 035103, (2002)
[31] M. E. J. Newman, S.Forrest, J.Balthrop, Email networks and the spread of computer viruses, Physical Review E 66,035101, (2002)
[32] W.Yu, S.Chellappan, X.Wang, D.Xuan, Peer-to-peer system-based active worm attacks: Modeling,analysis and defense. Computer Communications, v 31, n17, p 4005-4017 (2008)
[33] http://www.gnu.org/software/octave/
[34] G.Yan, S.Eidenbenz, Modeling Propagation Dynamics of Bluetooth Worms (Extended Version), IEEE Transactions on Mobile Computing, (2009)
[35] M.R.Faghani, H.Saidi, Malware Propagation in Online Social Networks. In proceeding of the 4th IEEE International malicious and unwanted programs (Malware09), Montreal, Canada (2009)
[36] M.R.Faghani, H.Saidi, Social networks' XSS worms, In proceeding of the 12th IEEE International conference on computational science and engineering (CSE09), Vancouver, Canada (2009)

[37] P.Holme, J.Beom, Growing scale-free networks with tunable clustering, Phys. Rev. E 65, pp. 026107-1:4 (2002)
[38] F.Viger, F.Latapy, Efficient and Simple Generation of Random Simple Connected Graphs with Prescribed Degree Sequence, Lecture notes in computer science, Vol. 3595, pp. 440-449.
[39] M.Thelwall, Social networks, gender, and friending: An analysis of mySpace member profiles, Journal of the American Society for Information Science and Technology, v 59, n 8, (2008), pp1321-1330.
[40] R.M.Anderson,The Samy Worm, “I'll never get caught. I'm Popular.”.
[41] Y.Ahn, S.Han, H.Kwak, S.Moon, H.Jeong, Analysis of Topological Characteristics of Huge Online Social Networking Services, Session:Semantic Web and Web 2.0 , WWW 2007,IW3C2, (2007)