



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΕΜΠΙΣΤΟΣΥΝΗ ΣΕ ΠΟΛΥΠΡΑΚΤΟΡΙΚΑ
ΣΥΣΤΗΜΑΤΑ**

Διπλωματική Εργασία

Ντόβας Αθανάσιος

Επιβλέποντες Καθηγητές:

Δασκαλοπούλου Ασπασία

Σταμούλης Γεώργιος

Βόλος 2019



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΕΜΠΙΣΤΟΣΥΝΗ ΣΕ ΠΟΛΥΠΡΑΚΤΟΡΙΚΑ
ΣΥΣΤΗΜΑΤΑ**

Διπλωματική Εργασία

Ντόβας Αθανάσιος

Επιβλέποντες Καθηγητές:

Δασκαλοπούλου Ασπασία

Σταμούλης Γεώργιος

Βόλος 2019



UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

TRUST IN MULTI-AGENT SYSTEMS

Diploma Thesis

Ntovas Athanasios

Supervisors:

Daskalopoulou Aspasia

Stamoulis Georgios

Volos 2019

ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της παρούσας διπλωματικής εργασίας, πρώτα απ' όλα θα ήθελα να ευχαριστήσω τους επιβλέποντες καθηγητές μου και ιδιαίτερα την Κυρία Ασπασία Δασκαλοπούλου για την εμπιστοσύνη που επέδειξε στο πρόσωπο μου, για την άριστη συνεργασία που είχαμε και για τις ουσιώδεις παρατηρήσεις και διορθώσεις καθ' όλη τη διάρκεια εκπόνησης της διπλωματικής εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω τους συμφοιτητές μου πρώτα απ' όλα για τη φιλία που αναπτύχτηκε μεταξύ μας και έπειτα για τη συνεργασία που είχαμε όλα αυτά τα χρόνια.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την υποστήριξη που μου παρείχε κατά τη διάρκεια των σπουδών μου.

ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας διπλωματικής εργασίας είναι η παρουσίαση και η ανάλυση των πιο σημαντικών τεχνικών υπολογισμού εμπιστοσύνης που μέχρι τώρα έχουν αναπτυχθεί στα πολυπρακτορικά συστήματα, καθώς επίσης γίνεται και μία σύντομη αναφορά στην έννοια της εξαπάτησης μεταξύ των πρακτόρων, μιας έννοιας που σχετικά πρόσφατα παρουσιάστηκε στην βιβλιογραφική έρευνα. Πιο συγκεκριμένα, βασικό κίνητρο αυτής της έρευνας αποτέλεσε η ανάγκη για τη συγκέντρωση και κατηγοριοποίηση των μοντέλων εμπιστοσύνης, μέσω των οποίων η μελέτη και η κατανόηση των μοντέλων αυτών γίνεται μια εύκολη υπόθεση. Κλείνοντας αυτή την εισαγωγή, να αναφερθεί ότι στο κεφάλαιο 2 γίνεται μια διάκριση των μοντέλων εμπιστοσύνης. Ακολούθως, παρουσιάζονται αναλυτικά τα μοντέλα εμπιστοσύνης, όπου για κάθε ένα μοντέλο μέσω μαθηματικών υπολογισμών εξάγεται τελικά ο τύπος υπολογισμού της εμπιστοσύνης. Στο κεφάλαιο 3 παρουσιάζεται η έννοια της εξαπάτησης, όπου γίνεται μια διάκριση στα είδη εξαπάτησης και αναφέρεται ένα μοντέλο εντοπισμού της εξαπάτησης.

ABSTRACT

The purpose for this thesis is the presentation and analysis of the most important Trust calculation techniques, which have been developed in Multi-agent Systems. The thesis also presents a short report of the meaning of deception within agents, a concept that was recently presented in the bibliography research. More precisely, the basic motive for this research was the need for gathering and categorization of trust models, through which the study and understanding of the models becomes an easy task. Closing this summary, it should be mentioned that chapter 2 presents a differentiation of trust models. Subsequently, there's an analysis of the trust models, where through mathematical calculations for each model we extract the type of trust calculation. In chapter 3, the meaning of deception is presented, where there's a division of the deception types and a model of deception detection is reported.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	v
ABSTRACT	v
ΠΕΡΙΕΧΟΜΕΝΑ.....	vi
ΚΕΦΑΛΑΙΟ 1	1
ΕΙΣΑΓΩΓΗ	1
ΚΕΦΑΛΑΙΟ 2	3
ΚΑΤΗΓΟΡΙΕΣ ΚΑΙ ΜΟΝΤΕΛΑ ΕΜΠΙΣΤΟΣΥΝΗΣ	3
2.1 Κατηγοριοποίηση Μοντέλων Εμπιστοσύνης	3
2.2 Μοντέλα εμπιστοσύνης e-commerce: eBay – Amazon	5
2.2.1 Διαδικασία ανατροφοδότησης eBay.....	6
2.2.2 Διαδικασία ανατροφοδότησης Amazon	7
2.2.3 Συμπεράσματα ανατροφοδότησης Amazon – eBay	8
2.3 Το Μοντέλο FIRE.....	9
2.3.1 Εμπιστοσύνη αλληλεπίδρασης (interaction trust).....	9
2.3.2 Εμπιστοσύνη βασισμένη στο ρόλο του πράκτορα (role based trust).....	10
2.3.3 Εμπιστοσύνη βασισμένη στη φήμη(Witness Reputation)	11
2.3.4 Φήμη πιστοποίησης (Certified Reputation).....	12
2.3.5 Υπολογισμός εμπιστοσύνης μοντέλου FIRE.....	12
2.3.6 Συμπεράσματα μοντέλου FIRE.....	13
2.4 Μοντέλο Συνδυαστικής Εμπιστοσύνης	14
2.4.1 Διάκριση εμπιστοσύνης	14
2.4.2 Υπολογισμός εμπιστοσύνης	14
2.4.3 Σύνολο E – Άμεση εμπιστοσύνη.....	15
2.4.4 Σύνολο R – Έμμεση εμπιστοσύνη.....	16
2.4.5 Σύνολο T – Συνολική Εμπιστοσύνη	16
2.4.6 Αλγόριθμος ανανέωσης εμπιστοσύνης	17
2.4.7 Συμπεράσματα	17
2.5 Το μοντέλο του Marsh.....	18
2.5.1 Υποθέσεις-Παραδοχές.....	18
2.5.2 Κατηγορίες και αναπαράσταση εμπιστοσύνης	18
2.5.3 Υπολογισμός εμπιστοσύνης	20
2.5.4 Κατώφλι συνεργασίας.....	21

2.5.5 Μνήμη	23
2.5.6 Ενημέρωση εμπιστοσύνης	23
2.5.7 Συμπεράσματα	24
2.6 Το Μοντέλο των L.Xiong και L.Liu (PeerTrust)	25
2.6.1 Εισαγωγή σε ένα Peer-to-Peer σύστημα	25
2.6.2 Υπολογισμός εμπιστοσύνης	26
2.6.3 Συμπεράσματα	29
2.7 Το μοντέλο των Mui-Mohtashemi-Halberstadt	30
2.7.1 Ορισμοί εννοιών.....	30
2.7.2 Υπολογισμός εμπιστοσύνης	32
2.7.3 Η διάδοση της φήμης.....	34
2.7.4 Συμπεράσματα	36
ΚΕΦΑΛΑΙΟ 3	37
ΕΞΑΠΑΤΗΣΗ ΚΑΙ ΕΝΤΟΠΙΣΜΟΣ ΕΞΑΠΑΤΗΣΗΣ ΣΤΑ ΠΟΛΥΠΡΑΚΤΟΡΙΚΑ ΣΥΣΤΗΜΑΤΑ	37
3.1 Εισαγωγή	37
3.2 Κατηγορίες εξαπάτησης	37
3.2.1 Κατηγοριοποίηση με βάση την φύση των συμμετεχόντων	37
3.2.2 Κατηγοριοποίηση με βάση την φύση της εξαπάτησης.....	39
3.3 Εντοπισμός εξαπάτησης.....	40
3.3.1 Θεωρητικό μοντέλο Εντοπισμού εξαπάτησης Johnson & Grazioli	41
3.3.2 Ολοκληρωμένο μοντέλο εντοπισμού εξαπάτησης	41
3.4 Συμπεράσματα	44
ΚΕΦΑΛΑΙΟ 4	46
ΣΥΜΠΕΡΑΣΜΑΤΑ	46
ΒΙΒΛΙΟΓΡΑΦΙΑ	48

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Κάποια μέρα αγοράζουμε μια καινούρια μοτοσυκλέτα. Το επόμενο βήμα για να αισθανόμαστε ασφαλείς είναι η αγορά ενός καλού κράνους που θα μας σώσει τη ζωή σε περίπτωση ατυχήματος. Στην αναζήτηση μας για ένα κράνος, παρατηρούμε ότι υπάρχει μεγάλη γκάμα τιμών, από πολύ φτηνά μέχρι πολύ ακριβά. Ασυναίσθητα, θεωρούμε ότι ένα ακριβό κράνος θα είναι αυτό που θα μας προστατεύει περισσότερο. Η εντύπωση αυτή μας δημιουργείται χωρίς να έχουμε ειδική γνώση της διαδικασίας κατασκευής ενός κράνους. Επιλέγουμε λοιπόν να αγοράσουμε ένα ακριβό κράνος γιατί εμπιστευόμαστε ότι το συγκεκριμένο είναι πιο ασφαλές από ένα φτηνότερο. Όταν βγαίνουμε για να γευματίσουμε σε ένα εστιατόριο, εμπιστευόμαστε τόσο το μάγειρα ότι θα μαγειρέψει καλά το φαγητό, όσο και τα υλικά που χρησιμοποιεί είναι υγιεινά και δε θα μας βλάψουν. Και στις δύο παραπάνω περιπτώσεις εμπεριέχεται το στοιχείο της εμπιστοσύνης. Οι άνθρωποι, στην καθημερινότητα μας καλούμαστε να πάρουμε μια σειρά από αποφάσεις που έχουν να κάνουν με σοβαρά και λιγότερο σοβαρά ζητήματα που στηρίζονται πάνω στην εμπιστοσύνη. Η εμπιστοσύνη βρίσκεται παντού στην καθημερινή μας ζωή, και αποτελεί ίσως τον πιο σημαντικό παράγοντα στη λήψη των αποφάσεων. Εμπιστευόμαστε από τον γιατρό που θα μας εγχειρήσει μέχρι το μανάβη για τα φρούτα που θα μας πουλήσει. Ακόμη και στην οδήγηση, υποσυνείδητα εμπιστευόμαστε τους άλλους οδηγούς ότι έχουν γνώση και θα σεβαστούν τον ΚΟΚ. Η εμπιστοσύνη λοιπόν υπάρχει παντού στην κοινωνία.

Από κοινωνιολογικής σκοπιάς, η εμπιστοσύνη έχει μελετηθεί εδώ και πάρα πολλά χρόνια και μπορούμε να πούμε ότι η ύπαρξη μιας κοινωνίας είναι άρρηκτα συνδεδεμένη με την έννοια της εμπιστοσύνης, και κατ' επέκταση η έλλειψη αυτής αποτελεί ανασταλτικό παράγοντα στην ανάπτυξη μιας κοινωνίας [1,2].

Όπως ακριβώς οι κοινωνιολόγοι διαπίστωσαν την εμπιστοσύνη στις διαπροσωπικές σχέσεις των ανθρώπων, με τον ίδιο τρόπο οι επιστήμονες των υπολογιστών παρατήρησαν την έννοια αυτή στον ψηφιακό κόσμο. Έτσι, έγινε η προσπάθεια να μελετηθεί σε βάθος η εμπιστοσύνη, να αναλύσουν τον ρόλο, τη σημαντικότητα και τελικά να προσπαθήσουν να ποσοτικοποιήσουν αυτή την όχι και τόσο σαφή έννοια.

Για τον ορισμό του τι είναι εμπιστοσύνη, έχουν διατυπωθεί πολλοί ορισμοί εκ' των οποίων οι δύο επικρατέστεροι στην βιβλιογραφία είναι οι παρακάτω:

- **Ορισμός Gambetta[3]:** Η εμπιστοσύνη ορίζεται ως ένα αποτέλεσμα συμπεριφοράς που βασίζεται σε επαρκείς θετικές προσδοκίες, οι οποίες είναι ικανές στο να μας κάνουν να εκτελέσουμε μια δράση.
- **Ορισμός Grandison και Sloman[4]:** Εμπιστοσύνη είναι η ισχυρή πεποίθηση στην ικανότητα μιας οντότητας ότι μπορεί να ενεργήσει αξιόπιστα και με ασφάλεια σε μια συγκεκριμένη περίπτωση.

Οι πράκτορες των πολυπρακτορικών συστημάτων, όπως ακριβώς και οι άνθρωποι δρουν αυτόνομα και πολλές φορές καλούνται να πάρουν αποφάσεις οι οποίες εμπεριέχουν ένα ποσοστό ρίσκου, όπως για παράδειγμα οι πράκτορες που συμμετέχουν σε μια αγοραπωλησία στο ηλεκτρονικό εμπόριο. Για το λόγο αυτό οι επιστήμονες που ασχολούνται με το τομέα αυτό όλα αυτά τα χρόνια έχουν δώσει μεγάλη βάση στην εμπιστοσύνη. Η επιστημονική μελέτη που έχει γίνει αποσκοπεί στο να δημιουργηθούν μοντέλα εμπιστοσύνης. Μοντέλα δηλαδή που στην ουσία προσπαθούν μέσω υπολογισμών να καταλήξουν σε μια τιμή για την εμπιστοσύνη, λαμβάνοντας υπόψη τα ιδιαίτερα χαρακτηριστικά που κάθε διαφορετικό πολυπρακτορικό σύστημα έχει, καθώς επίσης και τη φύση της κάθε συναλλαγής.

Στον αντίποδα της έννοιας της εμπιστοσύνης, συναντάμε την έννοια της εξαπάτησης. Συχνά, η εξαπάτηση οδηγεί στην έλλειψη εμπιστοσύνης. Με τον όρο εξαπάτηση εννοούμε την κάθε ενέργεια απόκρυψης της αλήθειας με σκοπό να επωφεληθούμε απ' αυτό. Σχετικά πρόσφατα η έννοια της εξαπάτησης έχει απασχολήσει τους ερευνητές, καθώς στη βιβλιογραφία είναι πολύ λίγες οι αναφορές που σχετίζονται με αυτή. Με την εισαγωγή των πρακτόρων σε όλες τις πτυχές της ζωής μας, η εξαπάτηση γίνεται ένα αντικείμενο έρευνας με ολοένα μεγαλύτερο ενδιαφέρον, καθώς τα συστήματα γίνονται όλο και πιο ανοιχτά.

Δεδομένης λοιπόν της ύπαρξης της εξαπάτησης στα πολυπρακτορικά συστήματα, το ζητούμενο είναι να γίνει αντιληπτή και να περιοριστεί. Ο εντοπισμός της εξαπάτησης, σε αντίθεση με τον υπολογισμό της εμπιστοσύνης είναι πιο απαιτητική δουλειά.

Οι έννοιες της εξαπάτησης και της εμπιστοσύνης παίζουν σημαντικό ρόλο στην ανάπτυξη των ψηφιακών κοινωνιών. Οι πράκτορες έχουν έρθει τόσο κοντά στην ανθρώπινη συμπεριφορά, που πολλές φορές η διάκριση γίνεται δύσκολα. Για το λόγο αυτό, οι σχέσεις των πρακτόρων μεταξύ τους, αλλά και μεταξύ ανθρώπων με πράκτορες είναι πολύ σημαντικό να διακατέχονται από εμπιστοσύνη, ώστε η συνεργασία να είναι εποικοδομητική και για τα δύο συναλλασσόμενα μέλη.

ΚΕΦΑΛΑΙΟ 2

ΚΑΤΗΓΟΡΙΕΣ ΚΑΙ ΜΟΝΤΕΛΑ ΕΜΠΙΣΤΟΣΥΝΗΣ

Η έννοια της εμπιστοσύνης στον ψηφιακό κόσμο οδήγησε στην ανάπτυξη κατάλληλων μοντέλων που επιτρέπουν τη διαχείρισή της. Από τα υπάρχοντα μοντέλα εμπιστοσύνης κάποια ακολουθούν απλή λογική υλοποίησης, ενώ κάποια άλλα χρησιμοποιούν πιο σύνθετες δομές και πολύπλοκους υπολογισμούς. Η διαβάθμιση αυτή έχει να κάνει με το περιβάλλον λειτουργίας, τις σχέσεις των πρακτόρων και το επίπεδο ασφάλειας που θέλουμε να πετύχουμε σε μια κοινωνία πρακτόρων. Στο κεφάλαιο αυτό θα γίνει παρουσίαση των μοντέλων εμπιστοσύνης που έχουν αναπτυχθεί και χρησιμοποιούνται σε πληθώρα εφαρμογών.

2.1 Κατηγοριοποίηση Μοντέλων Εμπιστοσύνης

Για την ευκολότερη μελέτη και κατανόηση τους, τα μοντέλα εμπιστοσύνης έχουν κατηγοριοποιηθεί ανάλογα με τον τρόπο υπολογισμού της συνάρτησης εμπιστοσύνης. Ένα μοντέλο εμπιστοσύνης συνήθως ανήκει σε περισσότερες από μια εκ των παρακάτω κατηγοριών λόγω της φύσης των σύγχρονων εφαρμογών που αναπτύσσονται. Οι κατηγορίες των μοντέλων εμπιστοσύνης είναι οι ακόλουθες[5,6]:

- **Μοντέλο ταυτότητας – Μοντέλο συμπεριφοράς**

Η διάκριση αυτών των μοντέλων γίνεται με κριτήριο τον τρόπο αξιολόγησης της εμπιστοσύνης. Στο **μοντέλο ταυτότητας** πρωτεύοντα ρόλο παίζει η ταυτότητα του πράκτορα. Έτσι ο πράκτορας τυγχάνει εμπιστοσύνης στην κοινότητα πρακτόρων αρκεί να δηλώσει την ταυτότητα του μέσω ενός συνθηματικού (π.χ. ενός κωδικού) που υποδηλώνει την ταυτότητα του πράκτορα στα άλλα μέλη. Στο **μοντέλο συμπεριφοράς** από την άλλη, για να γίνει κάποιος έμπιστο μέλος της κοινωνίας πρακτόρων αρκεί η συμπεριφορά που έχει επιδείξει στις προηγούμενες συναλλαγές να είναι θετική. Όπως θα εξετάσουμε, τα περισσότερα μοντέλα εμπιστοσύνης ανήκουν στην κατηγορία αυτή.

- **Μοντέλο κεντροποιημένο – Μοντέλο κατανεμημένο**

Η κατηγοριοποίηση αυτή βασίζεται στον τρόπο που η κοινωνία πρακτόρων διαχειρίζεται την εμπιστοσύνη. Στο **κεντροποιημένο μοντέλο** όλες οι αξιολογήσεις των πρακτόρων υπολογίζονται από μια κεντρική διαχείριση.

Δηλαδή, όταν ένας πράκτορας επιθυμεί να ενημερωθεί για την αξιολόγηση ενός άλλου πράκτορα απευθύνεται στον κεντρικό κόμβο. Στο **κατανεμημένο μοντέλο**, ο κάθε πράκτορας υπολογίζει μόνος του τη συνάρτηση εμπιστοσύνης για κάθε πράκτορα με τον οποίο συναλλάσσεται. Τα πιο διαδεδομένα μοντέλα εμπιστοσύνης χρησιμοποιούν την κατανεμημένη τεχνική υπολογισμού.

- **Μοντέλο αντικειμενικό– Μοντέλο υποκειμενικό**

Η διάκριση αυτή βασίζεται στην υποκειμενικότητα με την οποία υπολογίζεται η εμπιστοσύνη. Στο **αντικειμενικό μοντέλο** η εμπιστοσύνη προς έναν πράκτορα από όλα τα υπόλοιπα μέλη της κοινωνίας πρακτόρων θα είναι η ίδια ακριβώς, καθώς τα κριτήρια αξιολόγησης της συνάρτησης εμπιστοσύνης είναι σαφώς ορισμένα και ίδια για όλους τους πράκτορες. Ένα παράδειγμα από την καθημερινή ζωή είναι η εμπιστοσύνη που έχουμε στο manual ενός αυτοκινήτου ως προς τα τεχνικά χαρακτηριστικά του. Κανείς δεν αμφισβητεί τις πληροφορίες αυτές. Στην αντίθετη πλευρά, το **υποκειμενικό μοντέλο** οι πράκτορες δύναται να έχουν όλοι διαφορετική άποψη για έναν συγκεκριμένο πράκτορα. Ένα χαρακτηριστικό παράδειγμα είναι η αξιολόγηση μιας ταινίας σε ένα site ταινιών. Οι περισσότεροι έχουν διαφορετική άποψη για την ποιότητα της ταινίας γιατί δεν έχουν όλοι τις ίδιες προτιμήσεις. Άρα, εμπεριέχεται το στοιχείο της υποκειμενικότητας.

- **Μοντέλο συναλλαγών– Μοντέλο φήμης**

Τα μοντέλα αυτά βασίζονται στα δεδομένα που χρησιμοποιούνται στην αξιολόγηση της εμπιστοσύνης. Στο **μοντέλο συναλλαγών** λαμβάνεται υπόψη μόνο η προσωπική εμπειρία του πράκτορα με τον υπό εξέταση πράκτορα. Αντίθετα, στο **μοντέλο φήμης** λαμβάνονται υπόψη οι συναλλαγές όλων των πρακτόρων σε μια συνολική αντιπροσωπευτική άποψη. Επίσης, στο μοντέλο φήμης μπορεί να γίνει η διάκριση σε ενδογενές και εξωγενές. Στην ενδογενή φήμη η τιμή της συνάρτησης εμπιστοσύνης εξαρτάται περισσότερο από τις μαρτυρίες πρακτόρων οι οποίοι είναι φιλικοί προς εμάς. Στα εξωγενή συστήματα σε αντίθεση με τα ενδογενή, συμμετέχει στους υπολογισμούς της συνάρτησης εμπιστοσύνης και η κριτική πρακτόρων που είναι άγνωστοι προς εμάς. Τα σύγχρονα μοντέλα εμπιστοσύνης χρησιμοποιούν το μοντέλο φήμης.

- **Μοντέλο καθολικής πληροφορίας – Μοντέλο τοπικής πληροφορίας**

Τα μοντέλα εμπιστοσύνης μπορούν επίσης να διαχωριστούν σύμφωνα με τον τρόπο που συλλέγεται η πληροφορία για τον κάθε πράκτορα. Στο **μοντέλο καθολικής πληροφορίας** ένας πράκτορας μπορεί να έχει πρόσβαση στα δεδομένα που αφορούν όλες τις άλλες συναλλαγές των

υπολοίπων πρακτόρων. Η προσέγγιση αυτή, είναι αντικειμενική γιατί όλοι οι πράκτορες έχουν πρόσβαση στα ίδια δεδομένα. Η άλλη προσέγγιση είναι εκείνη του **μοντέλου τοπικής πληροφορίας**, όπου στα μοντέλα αυτής της κατηγορίας τα δεδομένα που χρησιμοποιούνται στον υπολογισμό της εμπιστοσύνης είναι τοπικά, δηλαδή προέρχονται είτε από το ίδιο το μέλος που διενεργεί την αξιολόγηση (προσωπικές εμπειρίες συναλλαγών), είτε από γειτονικά μέλη με τα οποία διατηρεί σχέσεις συνεργασίας. Από τον ορισμό αυτό, εύκολα καταλαβαίνουμε ότι το μοντέλο αυτό είναι υποκειμενικό καθώς η κρίση του κάθε πράκτορα έχει το στοιχείο της υποκειμενικότητας. Όσον αφορά το κόστος υπολογισμού της αξιολόγησης είναι πολύ μεγαλύτερο στην πρώτη περίπτωση και γι' αυτό στα καταναμεμημένα περιβάλλοντα εργασίας χρησιμοποιείται το μοντέλο τοπικών δεδομένων. Ωστόσο, το καθολικό μοντέλο υπολογισμού είναι πιο ακριβές γιατί έχει πρόσβαση σε όλα τα δεδομένα συναλλαγών.

- **Μοντέλο κατάταξης - Μοντέλο κατωφλίου**

Στο **μοντέλο κατάταξης**, το αποτέλεσμα της συνάρτησης δε χρησιμοποιείται για άμεση απόφαση συνεργασίας όπως στα άλλα μοντέλα, άλλα, η τιμή της συγκρίνεται με τις τιμές της αξιοπιστίας των υπολοίπων πρακτόρων και αναλόγως λαμβάνεται η απόφαση συνεργασίας ή μη συνεργασίας. Δηλαδή, από μόνη της μια τιμή αξιολόγησης δεν έχει αξία στο μοντέλο αυτό. Στο **μοντέλο κατωφλίου** ορίζεται μια τιμή για την συνάρτηση εμπιστοσύνης κάτω από την οποία δεν οδηγούμαστε σε συνεργασία των δυο πρακτόρων.

Η κατάταξη ενός μοντέλου εμπιστοσύνης σε μια από τις παραπάνω κατηγορίες, τις περισσότερες φορές είναι ξεκάθαρη και εξαρτάται από την φύση του μοντέλου.

2.2 Μοντέλα εμπιστοσύνης e-commerce: eBay – Amazon

Το eBay και το Amazon είναι οι δύο μεγαλύτερες πλατφόρμες ηλεκτρονικού εμπορίου σε παγκόσμιο επίπεδο, καθεμία από τις οποίες προσφέρει τη δυνατότητα αγοραπωλησίας προϊόντων μεταξύ δύο οποιονδήποτε συμβαλλόμενων μελών (αγοραστή και πωλητή). Οι αγορές μπορούν να λάβουν χώρα σε οποιαδήποτε σημεία του πλανήτη. Αρχικά, οι χρήστες δημιουργούν ένα προφίλ μέσω εγγραφής και έπειτα είναι σε θέση να συμμετέχουν σε δημοπρασίες και αγορές προϊόντων είτε ως αγοραστές είτε ως πωλητές. Αφού η συμφωνία συναλλαγής επιτευχθεί, ο αγοραστής καταθέτει τα χρήματα στον λογαριασμό του πωλητή και εκείνος με τη σειρά του αποστέλλει το προϊόν στον προορισμό του. Μετά την ολοκλήρωση της δημοπρασίας, ο αγοραστής έχει το δικαίωμα να αξιολογήσει τον πωλητή και το αντίστροφο. Συνεπώς, τα μοντέλα αυτά είναι μοντέλα φήμης. Μέχρι εδώ η διαδικασία φαντάζει απλή. Συνεπώς γεννάται το ερώτημα τι είναι αυτό που κάνει

αυτές τις πλατφόρμες να είναι τόσο δημοφιλείς; Μεταξύ άλλων πολύ σημαντικών λειτουργιών που προσφέρουν αυτές οι πλατφόρμες–διαμεσολαβητές όπως μάρκετινγκ και διαφήμιση, βασικό ρόλο παίζει η εμπιστοσύνη που έχουν τα συναλλασσόμενα μέλη μεταξύ τους. Δηλαδή, ότι το περιβάλλον συναλλαγών είναι αξιόπιστο (το ίδιο το σύστημα επιτρέπει ασφαλείς συναλλαγές) και σε περίπτωση μη τήρησης των συμφωνηθέντων, το ίδιο σύστημα είναι ικανό και έμπιστο να εφαρμόσει τις απαραίτητες κυρώσεις που προβλέπονται στους όρους χρήσης της πλατφόρμας. Η παραπάνω διαδικασία μας παραπέμπει σε κεντροποιημένο μοντέλο εμπιστοσύνης, καθώς όλες οι ενέργειες υπολογισμού της εμπιστοσύνης γίνονται από το σύστημα. Στη συνέχεια θα εξετάσουμε τα συστήματα ανατροφοδότησης αυτών των δύο συστημάτων.

2.2.1 Διαδικασία ανατροφοδότησης eBay

Το eBay διαθέτει ένα σύστημα ανατροφοδότησης το οποίο διέπεται από εξαιρετικά απλούς κανόνες. Αυτό το σύστημα ανατροφοδότησης στηρίζεται στην “online” φήμη που αναπτύσσεται σταδιακά με βάση τα σχόλια και τις αξιολογήσεις άλλων χρηστών. Υπάρχουν τέσσερα χαρακτηριστικά ανατροφοδότησης [7,8]:

- **Σκορ ανατροφοδότησης.** Αυτό το σκορ υπολογίζεται χρησιμοποιώντας ένα σύστημα αριθμητικής αξιολόγησης. Οι αγοραστές και οι πωλητές παίρνουν ένα βαθμό (+1) για μια θετική αξιολόγηση, μηδέν (0) για ουδέτερη αξιολόγηση και μείον ένα (-1) βαθμό για αρνητική αξιολόγηση. Αυτοί οι πόντοι αθροίζονται και βγαίνει το σκορ ανατροφοδότησης το οποίο αντιπροσωπεύεται από αστέρια. Για να ληφθεί υπόψη το σκορ ενός χρήστη πρέπει να λάβει δέκα θετικές αξιολογήσεις. Μέχρι τότε οποιαδήποτε αξιολόγηση που λαμβάνει ένας χρήστης δεν εμφανίζεται στους υπολοίπους χρήστες.
- **Ποσοστό ανατροφοδότησης.** Αυτό αναφέρεται στο ποσοστό των αγοραστών που αναφέρουν μια θετική εμπειρία από τη συναλλαγή με έναν πωλητή. Είναι το πιο σημαντικό χαρακτηριστικό γιατί μας δείχνει την αναλογία των θετικών ψήφων σε σχέση με τις συνολικές.
- **Σχόλια αξιολογήσεων.** Οι αγοραστές μπορούν να σχολιάσουν συγκεκριμένα σημεία της συναλλαγής όπως η ακρίβεια στην περιγραφή του προϊόντος, η επικοινωνία που είχε με τον πωλητή, ταχύτητα και τα έξοδα αποστολής κ.α.
- **Ομαδοποίηση αξιολογήσεων.** Συγκεκριμένα, οι αξιολογήσεις κάθε χρήστη ομαδοποιούνται κατά αποστολέα, και για κάθε αποστολέα υπολογίζεται το άθροισμα των τιμών αξιολόγησης. Έτσι για παράδειγμα εάν ένας αγοραστής έστειλε για ένα συγκεκριμένο πωλητή 2 θετικές και 3 αρνητικές αξιολογήσεις, το τελικό αποτέλεσμα θα είναι -1 που σημαίνει ότι έχει αρνητική άποψη για τον πωλητή. Το συνολικό άθροισμα όλων των χρηστών που έχουν αλληλεπιδράσει με έναν άλλο χρήστη με τον παραπάνω τρόπο αποτελεί και τον δείκτη αξιοπιστίας.

Τα προφίλ ανατροφοδότησης μπορούν να αναζητηθούν από το ID του κάθε μέλους. Τα προφίλ των πωλητών φαίνεται να είναι πιο καλά αξιολογημένα από αυτά των αγοραστών και αυτός είναι ένας λόγος για τον οποίο το Amazon δέχεται τον ένα τύπο αξιολόγησης έναντι του άλλου.

2.2.2 Διαδικασία ανατροφοδότησης Amazon

Σε γενικές γραμμές η περίπτωση του Amazon έχει πολλά κοινά χαρακτηριστικά με αυτή του eBay. Ωστόσο, σε ότι αφορά την ανατροφοδότηση του συστήματος υπάρχουν αρκετές διαφορές. Το Amazon είναι η μεγαλύτερη πλατφόρμα ηλεκτρονικού εμπορίου παγκοσμίως. Τα κύρια χαρακτηριστικά που διέπουν την πολιτική ανατροφοδότησης είναι τα εξής[9,8]:

- **Η ανατροφοδότηση διαφέρει από την κριτική.** Αν και παραπλήσιες έννοιες, διαφέρουν όσον αφορά η χρήση τους στην αξιολόγηση. Η ανατροφοδότηση στο Amazon σχετίζεται με την συμπεριφορά του πωλητή, κατά πόσο δηλαδή ο αγοραστής είναι ικανοποιημένος με τις υπηρεσίες που του παρείχε ο πωλητής όπως για παράδειγμα το χρόνο παράδοσης, τη συσκευασία, το επίπεδο επικοινωνίας κλπ. Από την άλλη μεριά, η κριτική έχει να κάνει με το κατά πόσο ο αγοραστής είναι ικανοποιημένος από αυτό καθ' αυτό το προϊόν που παρέλαβε (ποιότητα κατασκευής, λειτουργικότητα, γνησιότητα κλπ).
- **Αρνητική αξιολόγηση.** Η αξιολόγηση του πωλητή είναι μεγάλης σημασίας στο Amazon, ειδικά όταν είναι αρνητική. Η ίδια η πλατφόρμα μας συνιστά να υπάρχει όσο το δυνατό λιγότερη αρνητική αξιολόγηση. Υπάρχουν οδηγίες που προτείνουν μια αρνητική κριτική να γίνεται αφού πρώτα έχουν εξαντληθεί οι διαπραγματεύσεις μεταξύ αγοραστή – πωλητή προς την επίλυση οποιασδήποτε διαφωνίας. Έτσι προτείνει την αρνητική αξιολόγηση σαν τελευταίο μέσο μιας αρνητικής εμπειρίας.
- **Το σύστημα αξιολόγησης.** Το Amazon χρησιμοποιεί σύστημα πέντε αστέρων και είναι η καρδιά του συστήματος ανατροφοδότησης. Η διαβάθμιση γίνεται από το ένα αστέρι (πολύ κακή γνώμη) μέχρι τα πέντε αστέρια (πολύ καλή γνώμη):
 - Θετική ανατροφοδότηση 5 ή 4 αστέρια.
 - Ουδέτερη ανατροφοδότηση 3 αστέρια.
 - Αρνητική ανατροφοδότηση 2 ή 1 αστέρια.

Το σκορ της ανατροφοδότησης υπολογίζεται ως εξής :

$$FeedbackScore = \frac{Sum(Positive\ feedback)}{Sum(All\ feedback)}$$

Ένα παράδειγμα υπολογισμού του σκορ ανατροφοδότησης:

Ανατροφοδότηση	5	4	3	2	1
Αστέρια	55	35	5	3	2

Συνολικός αριθμός θετικών αξιολογήσεων = 90

Συνολικός αριθμός αξιολογήσεων = 100

Άρα Feedback score = 90%

- **Χρονικό περιθώριο αξιολογήσεων.** Οι αξιολογήσεις που έχει δεχθεί κάποιο μέλος υπάρχουν στο σύστημα για τους προηγούμενους 10 μήνες, καθώς και τα ποσοστά ανατροφοδότησης υπάρχουν για τις τελευταίες 30 , 90 , 365 μέρες και για όλη τη διάρκεια ζωής του μέλους στο σύστημα. Ο χρήστης έχει το περιθώριο 90 ημερών να αφήσει κάποιο σχόλιο και 60 ημερών (από τη στιγμή της δημοσιοποίησης) να το διαγράψει. Το Amazon έχει το δικαίωμα να διαγράψει σχόλια που παραβιάζουν τους όρους χρήσης του.

2.2.3 Συμπεράσματα ανατροφοδότησης Amazon – eBay

Η μέθοδος αξιολόγησης που ακολουθούν τα δύο αυτά μοντέλα είναι απλή και στηρίζεται στην ειλικρίνεια των χρηστών του. Υπάρχουν τα εξής θετικά σημεία στα συστήματα ανατροφοδότησης των δύο μοντέλων:

- Ο δείκτης αξιοπιστίας δημιουργεί ασφάλεια μεταξύ των αγοραστών και των πωλητών. Όσο μεγαλύτερος είναι ο δείκτης αυτός τόσο μεγαλύτερη και η εμπιστοσύνη σε μια συναλλαγή.
- Μέσω του δείκτη αξιοπιστίας μπορούν άγνωστοι μεταξύ τους χρήστες να σχηματίσουν άποψη ο ένας για τον άλλο.
- Δίνεται πολύ μεγάλη σημασία στην περίπτωση της αρνητικής αξιολόγησης. Ειδικά στην περίπτωση του Amazon αναφέρεται έντονα η όσο το δυνατό αποφυγή αρνητικής αξιολόγησης.
- Το μοντέλο του Amazon έχει διαβάθμιση (1-5) ως προς την ικανοποίηση που έλαβε ο αγοραστής από τον πωλητή και όχι μια απλή θετική/αρνητική αξιολόγηση όπως το μοντέλο του eBay.

Ωστόσο, υπάρχουν και αρνητικά σημεία:

- Πιθανή ύπαρξη αντιποίνων εξαιτίας μιας κακής αρχικής αξιολόγησης.
- Ένα κακό προφίλ κάποιου χρήστη μπορεί να αλλάξει απλά με την αλλαγή ταυτότητας.
- Κάθε εκτίμηση είναι υποκειμενική. Δηλαδή αυτό που για έναν χρήστη είναι αρνητικό για κάποιον άλλο μπορεί να μην είναι.

Παρά την ύπαρξη αρνητικών σχολίων, το μοντέλο του eBay και το μοντέλο του Amazon φαίνεται ότι δουλεύουν πολύ καλά. Τα συστήματα ανατροφοδότησης πάνω στα οποία στηρίζουν τη λειτουργία τους τα μοντέλα αυτά δημιουργούν

εμπιστοσύνη στους συναλλασσόμενους, πράγμα το οποίο αποδεικνύεται στην πράξη από τη χρήση των μοντέλων αυτών από εκατομμύρια χρήστες καθημερινά.

2.3 Το Μοντέλο FIRE

Το 2006 αναπτύχθηκε το μοντέλο εμπιστοσύνης FIRE [10] από τους T. Dong Huynh, Nicholas R. Jennings και Nigel R. Shadbolt. Το πεδίο εφαρμογής αυτού του μοντέλου είναι τα ανοιχτά πολυπρακτορικά συστήματα. Ως ανοιχτό ορίζουμε ένα πολυπρακτορικό σύστημα που ικανοποιεί μερικές προϋποθέσεις. Αρχικά, η εσωτερική αρχιτεκτονική κάθε πράκτορα να μην είναι δημόσια γνωστή στους άλλους πράκτορες. Δεύτερον, οι πράκτορες δεν νοιάζονται απαραίτητα για το κοινό όφελος και τρίτον, η συμπεριφορά των πρακτόρων δε μπορεί να προβλεφθεί. Οι πράκτορες σε ένα ανοιχτό πολυπρακτορικό σύστημα μπορούν να εισέλθουν και να εξέλθουν από το σύστημα οποιαδήποτε στιγμή.

Υποθέτουμε ότι στο σύστημα συμμετέχουν κάποιοι πράκτορες που αντιπροσωπεύουν το χρήστη και λειτουργούν ως πράκτορες-καταναλωτές, ενώ οι υπόλοιποι είναι αυτοί που παρέχουν κάποιες υπηρεσίες και λειτουργούν ως πράκτορες-πάροχοι. Το μοντέλο αυτό προσπαθεί να ενσωματώσει τέσσερα διαφορετικά είδη εμπιστοσύνης και φήμης για τον υπολογισμό ενός γενικού δείκτη αξιοπιστίας για τους πράκτορες που συμμετέχουν στο σύστημα αυτό και εξετάζονται.

2.3.1 Εμπιστοσύνη αλληλεπίδρασης (interaction trust)

Το πρώτο είδος είναι η εμπιστοσύνη αλληλεπίδρασης. Δηλαδή, η εμπιστοσύνη που προκύπτει από την απευθείας συναλλαγή των πρακτόρων. Έστω ότι ο πράκτορας-καταναλωτής a συναλλάσσεται με τον πράκτορα-πάροχο b σε μία συναλλαγή. Από τη συναλλαγή αυτή ο πράκτορας a θα αξιολογήσει τον πράκτορα b σχετικά με τις υπηρεσίες που του παρείχε (ταχύτητα εξυπηρέτησης, συνεργασία κλπ). Οι αξιολογήσεις παίρνουν τη μορφή $r = (a, b, i, c, v)$, όπου a και b οι πράκτορες που αναφέραμε πιο πριν που συμμετέχουν στη συναλλαγή i , v είναι η αξιολόγηση που δόθηκε από τον a στον b για την παρεχόμενη υπηρεσία c . Όπου $v \in [-1, 1]$. Αν $v = -1$ η αξιολόγηση είναι απολύτως αρνητική, αν $v = 1$ η αξιολόγηση είναι απολύτως θετική και αν $v = 0$ η αξιολόγηση είναι ουδέτερη. Το σύνολο των παλιότερων αξιολογήσεων του a προς τον b αναπαριστάται ως $R(a, b, c)$. Η εμπιστοσύνη αλληλεπίδρασης(IT) υπολογίζεται ως εξής:

$$T_I(a, b, c) = \sum_{r_i \in R(a, b, c)} \omega(r_i) \cdot v_i$$

Όπου v_i τιμή της αξιολόγησης r_i και $\omega(r_i)$ το βάρος της αξιολόγησης r_i . Για το βάρος $\omega(r_i)$ ισχύει ότι $\sum_{r_i \in R(a,b,c)} \omega(r_i) = 1$.

Στο μοντέλο FIRE, κάθε τιμή της εμπιστοσύνης ακολουθείται από ένα μέτρο αξιοπιστίας (ρ_T) που δείχνει κατά πόσο το μοντέλο παράγει αποτελέσματα που ανταποκρίνονται στην πραγματικότητα βάσει των τιμών που έχει λάβει σαν είσοδο. Το μέτρο της αξιοπιστίας προκύπτει από το γινόμενο των δύο ακολούθων τιμών:

- $\rho_N(a, b, c)$: το μέτρο της αξιοπιστίας που έχει να κάνει με τον αριθμό των παλιότερων αξιολογήσεων για τον υπολογισμό του T_i . Καθώς ο αριθμός n των παλιότερων αξιολογήσεων αυξάνεται, ο βαθμός της αξιοπιστίας μεγαλώνει και φτάνει στη τιμή κατωφλίου m .

$$\rho_N(a, b, c) = \begin{cases} \frac{n}{m} & , \quad \text{αν } n \leq m \\ 1 & , \quad \text{αν } n > m \end{cases} \quad , \quad \rho_N(a, b, c) \in [0,1]$$

- $\rho_D(a, b, c)$: η αξιολόγηση της απόκλισης της αξιοπιστίας. Δηλαδή όσο πιο μεγάλη είναι η διασπορά των τιμών της αξιολόγησης, τόσο πιο απρόβλεπτος θεωρείται ο πράκτορας b στο να συνεργαστεί με τον πράκτορα a .

$$\rho_D(a, b, c) = 1 - \sum_{r_i \in R(a,b,c)} \frac{\omega(r_i) \cdot |v_i - T_i(a, b, c)|}{2}$$

$$\text{Άρα} \quad \rho_{T_i}(a, b, c) = \rho_D(a, b, c) \rho_N(a, b, c)$$

2.3.2 Εμπιστοσύνη βασισμένη στο ρόλο του πράκτορα (role based trust)

Οι ρόλοι που έχουν οι πράκτορες σε μια κοινωνία επηρεάζουν και το κατά πόσο μπορεί να εμπιστευτεί ο ένας πράκτορας τον άλλο. Για παράδειγμα, μεταξύ δύο πρακτόρων που ανήκουν στην ίδια ομάδα, αναπτύσσεται κλίμα εμπιστοσύνης ευκολότερα απ' ότι μεταξύ δυο πρακτόρων με αντικρουόμενα ή καθόλου συμφέροντα. Όμως, σε μια κοινωνία οι ρόλοι διαρκώς μεταβάλλονται εξαιτίας των προσωπικών συμφερόντων του κάθε μέλους. Συνεπώς είναι δύσκολο να αναπτυχθεί μια φόρμουλα υπολογισμού της εμπιστοσύνης στην περίπτωση αυτή. Για το λόγο αυτό, χρησιμοποιείται ένα σύστημα κανόνων για να προσδιορίσουμε την εμπιστοσύνη που βασίζεται στο ρόλο του πράκτορα (Role Trust). Το σύστημα κανόνων που χρησιμοποιείται έχει τη μορφή $rul = (role_a, role_b, c, v_D, e_D)$. Για να γίνουμε πιο σαφείς θα εξηγήσουμε τι σημαίνει αυτή η σχέση. Όπου $role_a$ και $role_b$ οι ρόλοι που έχουν οι πράκτορες a και b αντίστοιχα, c είναι το πεδίο για το οποίο αξιολογείται ο πράκτορας b από τον πράκτορα a . Η αναμενόμενη συμπεριφορά του b στην αλληλεπίδραση με τον a είναι v_D με $v_D \in [-1,1]$. e_D είναι το προκαθορισμένο επίπεδο επίδρασης του κανόνα αυτού στο αποτέλεσμα της τιμής εμπιστοσύνης. Ισχύει ότι $e_D \in [0,1]$.

Όλα τα παραπάνω γίνονται πιο ξεκάθαρα με ένα παράδειγμα. Έστω ότι συμβαίνει μια αγοραπωλησία. Στην αγοραπωλησία αυτή λαμβάνει μέρος ο αγοραστής ($role_a$) και ο πωλητής($role_b$) και το πεδίο αξιολόγησης(c) είναι η ποιότητα του προϊόντος ή της υπηρεσίας. Συμφώνα με τον κανόνα που έχουμε περιγράψει, μια περίπτωση κανόνα στο συγκεκριμένο παράδειγμα θα μπορούσε να είναι η εξής: $rul_1 = (seller, buyer, quality, -0.2, 0.3)$. Η περίπτωση αυτή έχει να κάνει με άγνωστους πράκτορες και το προϊόν είναι κακής ποιότητας ($c = -0.2$). Μια άλλη περίπτωση θα ήταν αν η αγοραπωλησία λάμβανε μέρος μεταξύ φίλων και το προϊόν/υπηρεσία είναι αυτό ακριβώς που είχε συμφωνηθεί($c=0$). $rul_2 = (friend_seller, friend_buyer, quality, 0.6, 0.8)$

Κάθε πράκτορας έχει τους δικούς του κανόνες που αποθηκεύονται σε μία δική του βάση δεδομένων. Η τιμή της RT δίνεται από την σχέση:

$$T_R(a, b, c) = \frac{\sum_{rul_i \in Rules(a, b, c)} e_{Di} \cdot v_i}{\sum_{rul_i \in Rules(a, b, c)} e_{Di}} \quad (5)$$

Όπου $rul_i = (role_a, role_b, c, v_{Di}, e_{Di})$ είναι ένα σετ κανόνων που ανήκει στο $Rules(a, b, c)$,με $Rules(a, b, c)$ είναι ένα υποσύνολο κανόνων που περιλαμβάνει τους κανόνες των δύο πρακτόρων και τους όρους c .

Η αξιοπιστία υπολογίζεται ακριβώς όπως στο 2.3.1 και γράφεται ως : $\rho_{T_R}(a, b, c)$

2.3.3 Εμπιστοσύνη βασισμένη στη φήμη (Witness Reputation)

Η εμπιστοσύνη που βασίζεται στη φήμη είναι η έμμεση εμπιστοσύνη που έχει να κάνει με τις αξιολογήσεις των αλληλεπιδράσεων που είχαν οι άλλοι πράκτορες με τον υπό-αξιολόγηση πράκτορα. Όταν ο πράκτορας a θέλει να εκτιμήσει τη φήμη για τον πράκτορα b για ένα συγκεκριμένο χαρακτηριστικό c (αυτό ορίζεται ως $T_W(a, b, c)$), τότε στέλνει αιτήματα της μορφής $(_, b, _, c, _)$ σε όλους τους γνωστούς του πράκτορες που θεωρεί ότι πιθανόν να γνωρίζουν τον πράκτορα b για το χαρακτηριστικό c . Οι πράκτορες με τη σειρά τους συγκρίνουν το αίτημα με τις δικές τους βάσεις δεδομένων για τυχόν ταιριάσματα με το αίτημα . Αν δε βρουν κανένα ταιρίασμα, τότε επιστρέφουν στον a μια λίστα με πράκτορες που πιθανόν να μπορούν εκείνοι να απαντήσουν στα αιτήματα αυτά. Η διαδικασία συνεχίζεται έως ότου να βρεθεί κάποιος πράκτορας που να απαντάει στο αίτημα του a . Το πόσο μακριά είναι ο πράκτορας που απαντάει στο αίτημα του a ορίζει το μήκος της αλυσίδας. Η διαδικασία επίσης μπορεί να σταματήσει όταν το μήκος της αλυσίδας ξεπεράσει μια προκαθορισμένη τιμή, τη τιμή κατωφλίου. Η τιμή κατωφλίου ορίζεται γιατί όσο πιο μακριά βρεθεί πράκτορας που να ικανοποιεί το αίτημα του a , τόσο πιο αναξιόπιστος είναι.

Η εμπιστοσύνη που βασίζεται στη φήμη υπολογίζεται ως εξής:

$$T_w(a, b, c) = \sum_{r_i \in R_w(a, b, c)} \omega(r_i) v_i \quad (6)$$

Όπου $R_w(a, b, c)$ είναι το σύνολο των αξιολογήσεων που φτάνουν στον πράκτορα a , το βάρος $\omega(r_i)$ που ορίζεται για κάθε αξιολόγηση όπως στην παράγραφο 2.3.1 . v_i είναι η τιμή της αξιολόγησης r_i .

Η αξιοπιστία υπολογίζεται ακριβώς όπως στο 2.3.1 και γράφεται ως : $\rho_{T_w}(a, b, c)$

2.3.4 Φήμη πιστοποίησης (Certified Reputation)

Η τελευταία κατηγορία ονομάζεται φήμη πιστοποίησης και παρουσιάζει τους δείκτες αξιολόγησης ενός πράκτορα-παρόχου b , οι οποίοι προέκυψαν από άλλους πράκτορες που είχαν αλληλεπιδράσει μαζί του στο παρελθόν[11]. Από την στιγμή που ο πράκτορας b μπορεί να διαλέξει ο ίδιος ποιες αξιολογήσεις θα παρουσιάσει στον πράκτορα a , ένας τυχαίος πράκτορας είναι πιθανό να παρουσιάσει τις πιο καλές αξιολογήσεις. Γι' αυτό θεωρούμε ότι η φήμη αυτού του τύπου πιθανόν να υπερεκτίμα τη συμπεριφορά του πράκτορα b . Ωστόσο, το μεγάλο πλεονέκτημα αυτού του τύπου της εμπιστοσύνης είναι η άμεση διαθεσιμότητα της φήμης του πράκτορα που θέλουμε να αξιολογήσουμε. Ακόμη και αν οι άλλες μέθοδοι υπολογισμού της εμπιστοσύνης αποτύχουν, αυτή η μέθοδος θα δώσει αποτελέσματα αρκεί να έχουν προηγηθεί συναλλαγές για τον πράκτορα b .

Ο τύπος υπολογισμού της εμπιστοσύνης βασισμένος στη φήμη πιστοποίησης είναι ο ίδιος με την περίπτωση της εμπιστοσύνης βασισμένης στη φήμη, με τη διαφορά ότι το σύνολο των αξιολογήσεων αναφέρεται από τον πράκτορα b στον πράκτορα b .

$$T_w(a, b, c) = \sum_{r_i \in R_w(b, b, c)} \omega(r_i) v_i \quad (7)$$

Η αξιοπιστία υπολογίζεται ακριβώς όπως στο 2.3.1 και γράφεται ως : $\rho_{T_c}(a, b, c)$

2.3.5 Υπολογισμός εμπιστοσύνης μοντέλου FIRE

Στο μοντέλο αυτό, η εμπιστοσύνη υπολογίζεται βάσει των τεσσάρων κατηγοριών που αναλύθηκαν παραπάνω. Συγκεκριμένα, χρησιμοποιείται η μέθοδος του σταθμισμένου μέσου με διαφορετικά βάρη για τους δείκτες της κάθε κατηγορίας, ώστε να προκύψει ένας γενικός δείκτης που να αντιπροσωπεύει την αξιοπιστία του πράκτορα που προσφέρει τις υπηρεσίες του. Τα βάρη που χρησιμοποιούνται προκύπτουν από τον χρήστη και εξαρτώνται από τις προτιμήσεις του και την σπουδαιότητα της κάθε περίπτωσης. Η εμπιστοσύνη λοιπόν υπολογίζεται ως εξής:

$$T_{(a,b,c)} = \frac{\sum_{k \in \{I,R,W,C\}} w_k \cdot T_k(a, b, c)}{\sum_{k \in \{I,R,W,C\}} w_k}$$

Και το μέτρο της αξιοπιστίας του της τιμής της συνδυασμένης εμπιστοσύνης είναι:

$$\rho_T(a, b, c) = \frac{\sum_{k \in \{I,R,W,C\}} w_k}{\sum_{k \in \{I,R,W,C\}} W_k}$$

Όπου $w_k = W_k \cdot \rho_{T_k}(a, b, c)$ με W_k να παίρνει μία εκ των τιμών W_I, W_R, W_W και W_C που είναι τα βάρη που ορίζει ο χρήστης για την κάθε περίπτωση εμπιστοσύνης ξεχωριστά ανάλογα με τη σημαντικότητα της κάθε μιας. Όπου $\rho_{T_k}(a, b, c)$ είναι η αξιοπιστία της κάθε περίπτωσης εκ των τεσσάρων παραπάνω κατηγοριών. Τέλος, $T_k(a, b, c)$ είναι η εμπιστοσύνη για κάθε μια κατηγορία εμπιστοσύνης.

2.3.6 Συμπεράσματα μοντέλου FIRE

Το μοντέλο FIRE ακριβώς επειδή χρησιμοποιείται σε ανοιχτά πολυπρακτορικά συστήματα (open MAS) ανήκει στα κατανεμημένα μοντέλα. Το γεγονός ότι προσπαθεί να εκτιμήσει την εμπιστοσύνη για οποιοδήποτε πράκτορα μέσα στην κοινωνία των πρακτόρων, καθιστά τον υπολογισμό της εμπιστοσύνης δύσκολο και πολλές φορές ανακριβή, ειδικά όταν πρόκειται για νέο μέλος στην κοινωνία αυτή. Η πιθανότητα ένα νέο μέλος να έχει αλληλεπιδράσει με τον πράκτορα που αξιολογεί, σε μεγάλης κλίμακας κοινωνίες είναι ιδιαιτέρως μικρή, συνεπώς κάθε προσπάθεια αξιολόγησης μέσω άλλων πρακτόρων-μαρτύρων τείνει στην αποτυχία. Για κάποιο παλιό πράκτορα με πολλές αλληλεπιδράσεις τα πράγματα γίνονται πιο σίγουρα στην περίπτωση της εμπιστοσύνης βασισμένης στη φήμη. Όσον αφορά την περίπτωση της εμπιστοσύνης φήμης πιστοποίησης, η πίστη ότι ο πράκτορας που αξιολογείται θα μας δώσει αληθή αποτελέσματα των παλιότερων αξιολογήσεων που έχει λάβει παίζει τον πιο σημαντικό ρόλο. Η περίπτωση αυτή κάνει το μοντέλο αρκετά ευάλωτο. Οι δύο αυτές περιπτώσεις λαμβάνουν τα δεδομένα τους από αναφορές τρίτων και μπορούν να δώσουν ανακριβή και λανθασμένα αποτελέσματα, πράγμα το οποίο είναι αναπόφευκτο στα ανοιχτά πολυπρακτορικά συστήματα[12]. Από την άλλη, τα δύο πρώτα είδη που αναλύθηκαν, χρησιμοποιούν την απευθείας προσωπική εμπειρία ενός πράκτορα με τον πράκτορα που επιθυμεί να αξιολογήσει. Για το λόγο αυτό τα μοντέλα αυτά δίνουν πολύ καλά αποτελέσματα.

Η συνδυαστική φύση του μοντέλου FIRE το καθιστά ένα ισχυρό εργαλείο στα ανοιχτά πολυπρακτορικά συστήματα. Για το λόγο αυτό επίσης, το μοντέλο είναι κατάλληλο για περιβάλλοντα που αλλάζουν με γρήγορο ρυθμό παρ' όλα τα αρνητικά που είναι σίγουρο ότι θα υπάρξουν σε τέτοιου είδους συστήματα.

2.4 Μοντέλο Συνδυαστικής Εμπιστοσύνης

Το 2013 οι Manh Hung Nguyen και Dinh Que Tran εισήγαγαν ένα μοντέλο εμπιστοσύνης [13] που σκοπό είχε να παρουσιάσει ένα υβριδικό τρόπο υπολογισμού της εμπιστοσύνης συνδυάζοντας την άμεση και την έμμεση εμπιστοσύνη. Επίσης παρουσιάζεται ένας αλγόριθμος με τον τρόπο που ανανεώνεται η εμπιστοσύνη.

2.4.1 Διάκριση εμπιστοσύνης

Στο μοντέλο αυτό η εμπιστοσύνη χωρίζεται σε τρεις βασικές κατηγορίες:

- Εμπιστοσύνη βασισμένη στην προσωπική εμπειρία (άμεση), όπου όπως έχουμε ήδη δει είναι η εμπιστοσύνη που αποκτά ένας πράκτορας για τους άλλους πράκτορες μέσω των δικών του αλληλεπιδράσεων με αυτούς.
- Εμπιστοσύνη βασισμένη στις αναφορές άλλων πρακτόρων (έμμεση), όπου ο πράκτορας αξιολογεί τους άλλους πράκτορες βάσει των μαρτυριών τους για τον υπό-αξιολόγηση πράκτορα.
- Συνολική εμπιστοσύνη. Είναι η εμπιστοσύνη που βασίζεται τόσο στην άμεση εμπειρία, όσο και στην έμμεση. Τα μοντέλα που χρησιμοποιούν αυτού του είδους την αξιολόγηση εμπιστοσύνης ονομάζονται υβριδικά.

2.4.2 Υπολογισμός εμπιστοσύνης

Θεωρούμε ότι ένα σύστημα πρακτόρων στο οποίο η εμπιστοσύνη αναπαρίσταται από τα σύνολα (A, E, R, T) , εκ των οποίων:

- $A = \{1, 2, \dots, n\}$ είναι το σύνολο των πρακτόρων που συμμετέχουν στο σύστημα.
- $E = |A| * |A|$, ο πίνακας της άμεσης εμπιστοσύνης. Το στοιχείο E_{ij} αναπαριστά την άμεση εμπιστοσύνη που αποκτά ο πράκτορας i από τη συναλλαγή με τον πράκτορα j .
- $R = |A| * |A|$, ο πίνακας της έμμεσης εμπιστοσύνης. Το στοιχείο R_{ij} αναπαριστά την έμμεση εμπιστοσύνη που λαμβάνει ο πράκτορας i από άλλους πράκτορες για τον πράκτορα j .
- $T = |A| * |A|$, ο πίνακας της συνολικής εμπιστοσύνης. Το στοιχείο T_{ij} αναπαριστά τη συνολική εμπιστοσύνη του πράκτορα i προς τον πράκτορα j .

Ακολούθως θα παρουσιάσουμε τον υπολογισμό των παραπάνω συνόλων ξεχωριστά.

2.4.3 Σύνολο E – Άμεση εμπιστοσύνη

Έστω ότι το σύνολο U_{ij} αναπαριστά το πλήθος των συναλλαγών του πράκτορα i με τον πράκτορα j που έχουν πραγματοποιηθεί από την πρώτη συναλλαγή μέχρι και την πιο πρόσφατη. Οι συναλλαγές έχουν χρονολογική σειρά. Με k συμβολίζεται η k -οστη (k^{th}) συναλλαγή που έλαβε χώρα μεταξύ του i και του j .

Εμπιστοσύνη συναλλαγής t_{ij}^k

Η εμπιστοσύνη συναλλαγής t_{ij}^k , όπου $t_{ij}^k \in [0,1]$, $k = 1, 2, \dots, |U_{ij}|$ και $i, j = 1, 2, \dots, n$. Η εμπιστοσύνη συναλλαγής t_{ij}^k αναπαριστά την εμπιστοσύνη που έχει ο πράκτορας i στον πράκτορα j από την k -οστη συναλλαγή μεταξύ τους. Αν η τιμή της εμπιστοσύνης είναι 1 τότε η αξιολόγηση είναι θετική, αν είναι 0.5 είναι ουδέτερη και τέλος είναι 0 είναι αρνητική.

Διάνυσμα συναλλαγών εμπιστοσύνης t_{ij}

Το διάνυσμα t_{ij} περιέχει όλες τις τιμές της εμπιστοσύνης των συναλλαγών t_{ij}^k που έχουν πραγματοποιηθεί μεταξύ του i και του j πράκτορα.

Διάνυσμα βάρους συναλλαγών w

Διάνυσμα βάρους συναλλαγών $w = (w_1, w_2, \dots, w_{|U_{ij}|})^T$ είναι το διάνυσμα που περιέχει τα βάρη που ορίζουμε για την κάθε περίπτωση συναλλαγής. Ισχύει ότι $w_k \in [0,1]$, $k = 1, 2, \dots, |U_{ij}|$ είναι το βάρος για την k -οστη συναλλαγή δεδομένου ότι $w_{k_1} \geq w_{k_2}$ αν $k_1 < k_2$ και $\sum_{k=1}^{|U_{ij}|} w_k = 1$. Οι τιμές του διανύσματος w είναι μεγαλύτερες όσο πιο πρόσφατες είναι οι συναλλαγές και μικρότερες όσο πιο παλιές. Πρακτικά, αυτό σημαίνει ότι όσο πιο πρόσφατη είναι μια συναλλαγή, τόσο πιο σημαντική είναι για τον υπολογισμό της άμεσης εμπιστοσύνης.

Αν η συνάρτηση $Q: [0,1] \rightarrow [0,1]$ που ικανοποιεί τις συνθήκες $Q(0) = 1$, $Q(1) = 0$ και $Q(i_1) \geq Q(i_2)$, τότε αν $i_1 < i_2$ (φθίνουσα συνάρτηση), τότε το βάρος w_i παράγεται μέσω της Q με τον εξής τρόπο:

$$w_i = Q\left(\frac{i-1}{|U_{ij}|}\right) - Q\left(\frac{i}{|U_{ij}|}\right), \text{ για } i = 1, 2, \dots, |U_{ij}|$$

Άμεση εμπιστοσύνη E_{ij}

Η άμεση εμπιστοσύνη E_{ij} του πράκτορα i στον πράκτορα j ορίζεται ως:

$$E_{ij} = t_{ij} w = \sum_{k=1}^{|U_{ij}|} t_{ij}^k w_k$$

Το διάνυσμα E_i αναπαριστά την άμεση εμπιστοσύνη του πράκτορα i προς όλους τους άλλους πράκτορες και λέγεται διάνυσμα άμεσης εμπιστοσύνης.

2.4.4 Σύνολο R – Έμμεση εμπιστοσύνη

Έστω ότι ο πράκτορας i επιθυμεί να αλληλεπιδράσει με τον πράκτορα j και ότι δεν έχουν αλληλεπιδράσει ποτέ ξανά στο παρελθόν. Έστω ότι V_{ij} είναι το σύνολο των πρακτόρων με τους οποίους έχει αλληλεπιδράσει στο παρελθόν τόσο ο πράκτορας i όσο και ο πράκτορας j . Έστω $r_{ij}^l \in [0,1], l \in V_{ij}$ είναι η αξιολόγηση που λαμβάνει ο πράκτορας i για τον πράκτορα j μέσω του πράκτορα l .

Η εμπιστοσύνη αναφοράς (έμμεση εμπιστοσύνη) είναι μια συνάρτηση υπολογίζει τον μέσο όρο όλων των αξιολογήσεων, δηλαδή:

$$R_{ij} = \begin{cases} \sum_{l \in V_{ij}} \frac{r_{ij}^l}{|V_{ij}|} & , \text{ αν } V_{ij} \neq 0 \\ 0 & , \text{ αλλιώς} \end{cases}$$

Το διάνυσμα έμμεσης εμπιστοσύνης R_i αναπαριστά την εμπιστοσύνη αναφοράς του πράκτορα i προς όλους τους πράκτορες του συστήματος.

2.4.5 Σύνολο T – Συνολική Εμπιστοσύνη

Ο συνδυασμός του συνόλου R και του συνόλου E μας δίνει τη συνολική (συνδυαστική) εμπιστοσύνη $t(e, r) \in [0,1]$. Για τη συνολική εμπιστοσύνη πρέπει να ικανοποιούνται οι ακόλουθες συνθήκες:

- **$\min(e, r) \leq t(e, r) \leq \max(e, r)$**
Η συνδυαστική εμπιστοσύνη πρέπει να είναι μεταξύ του ελαχίστου και του μεγίστου της έμμεσης και της άμεσης εμπιστοσύνης.
- **$t(e_1, r) \leq t(e_2, r) \text{ , αν } e_1 \leq e_2$**
Όσο πιο μεγάλη είναι η άμεση εμπιστοσύνη , τόσο πιο μεγάλη θα είναι η συνδυαστική εμπιστοσύνη.
- **$t(e, r_1) \leq t(e, r_2) \text{ , αν } r_1 \leq r_2$**
Όσο πιο μεγάλη είναι έμμεση εμπιστοσύνη , τόσο πιο μεγάλη θα είναι η συνδυαστική εμπιστοσύνη.

Η συνδυαστική εμπιστοσύνη T_{ij} ενός πράκτορα i προς έναν άλλο πράκτορα j ορίζεται ως $T_{ij} = t(E_{ij}, R_{ij})$, δηλαδή ως συνδυασμός της άμεσης και της έμμεσης εμπιστοσύνης. Συγκεκριμένα υπολογίζεται ως εξής:

$$T_{ij} = W_{ie}E_{ij} + W_{ir}R_{ij}$$

Όπου W_{ie} και W_{ir} τα βάρη για τις E_{ij} και R_{ij} αντίστοιχα.

2.4.6 Αλγόριθμος ανανέωσης εμπιστοσύνης

Η εμπιστοσύνη του πράκτορα i στον πράκτορα j αλλάζει καθ' όλη τη διάρκεια της ζωής του i είτε επειδή αλλάζει η άμεση εμπιστοσύνη εξαιτίας μιας αλληλεπίδρασης με τον j , είτε επειδή αλλάζει η έμμεση εμπιστοσύνη εξαιτίας μιας μαρτυρίας ενός πράκτορα l που αλληλεπίδρασε με τον j . Ο αλγόριθμος που δίνει το μοντέλο για την ανανέωση της εμπιστοσύνης στις δύο παραπάνω περιπτώσεις είναι ο εξής:

```

1: for all agent  $i$  in the system do
2:   if (there is a new transaction  $k - th$  with agent  $j$ ) or (there is a new reference
      trust  $E_{lj}$  from agent  $l$  about agent  $j$ ) then
3:     if (there is a new transaction  $k - th$  with agent  $j$ ) then
4:        $t_{ij}^k \leftarrow 0 \text{ or } 0.5 \text{ or } 1$  //0:negative,0.5:neutral,1:positive
5:        $t_{ij} \leftarrow t_{ij} \cup t_{ij}^k$  //add  $t_{ij}^k$  into  $t_{ij}$ 
6:        $t_{ij} \leftarrow \text{Sort}(t_{ij})$  //sort  $t_{ij}$  by the time
7:        $w \leftarrow \text{Generate } W(k)$  //generate the weight vector  $w$ 
8:        $E_{ij} \leftarrow \sum_{h=1}^k t_{ij}^h * w_h$  //update the experience trust
9:     end if
10:    If (there is a new reference trust  $E_{lj}$  from agent  $l$  about agent  $j$ ) then
11:      if ( $l$  is a new friend) then
12:         $V_{ij} \leftarrow V_{ij} \cup \{l\}$  // add  $l$  into friends set  $V_{ij}$ 
13:      end if
14:       $r_{ij}^l \leftarrow E_{lj}$ 
15:       $R_{ij} \leftarrow \frac{\sum_{h \in V_{ij}} r_{ij}^h}{|V_{ij}|}$  //update the reference trust
16:    end if
17:     $T_{ij} = t(E_{ij}, R_{ij})$  //update the final trust
18:  end if
19 end for

```

2.4.7 Συμπεράσματα

Όπως όλα τα υβριδικά μοντέλα, το μοντέλο των Manh Hung Nguyen και Dinh Que Tran είναι ένα συνδυαστικό μοντέλο που χρησιμοποιεί στους υπολογισμούς του τόσο την προσωπική εμπειρία όσο και τη φήμη. Το μοντέλο αυτό έχει δοκιμαστεί σε περιβάλλοντα ηλεκτρονικού εμπορίου και τα αποτελέσματα έχουν δείξει ότι είναι προτιμότερο να χρησιμοποιούμε τον συνδυασμό της προσωπικής εμπειρίας και τις αναφορών άλλων πρακτόρων για την αξιολόγηση ενός πράκτορα. Τέλος, έχει γίνει η παραδοχή ότι όλοι οι πράκτορες είναι έμπιστοι ως προς τις πληροφορίες που

ανταλλάσσουν. Ωστόσο, υπάρχουν πράκτορες που διαδίδουν ψευδείς πληροφορίες και αυτό είναι ένα αρνητικό σημείο για το μοντέλο.

2.5 Το μοντέλο του Marsh

Ένα μοντέλο ορόσημο στη θεωρία της εμπιστοσύνης στα πολυπρακτορικά συστήματα είναι το μοντέλο που ανέπτυξε ο Marsh το 1994 [1]. Ο Marsh είναι από τους πρώτους που ασχολήθηκαν με τον μαθηματικό φορμαλισμό της έννοιας της εμπιστοσύνης. Επιχείρησε να αποτυπώσει με μια μαθηματική αναπαράσταση με σχετικά απλούς υπολογισμούς τη μέχρι τότε ασαφή για πολλούς έννοια της εμπιστοσύνης. Πέρα από το γεγονός ότι η εργασία του Marsh αποτέλεσε θεμέλιο στη θεωρία της εμπιστοσύνης, η σπουδαιότητα της αντικατοπτρίζεται και στο πλήθος των αναφορών στην εργασία αυτή από άλλους ερευνητές που ασχολήθηκαν με τον τομέα της εμπιστοσύνης.

2.5.1 Υποθέσεις-Παραδοχές

Το μοντέλο βασίζεται στις εξής παραδοχές μεταξύ των πρακτόρων [10] :

- Δεν υπάρχει το αίσθημα της υποχρέωσης όσον αναφορά τις συναλλαγές μεταξύ των πρακτόρων. Δηλαδή, ένας πράκτορας δεν είναι σίγουρο ότι θα αλληλεπιδράσει με κάποιον άλλον πράκτορα επειδή του οφείλει κάτι.
- Η αυτονομία ενός πράκτορα να αλληλεπιδρά ή όχι με έναν άλλον πράκτορα θεωρείται δεδομένη.
- Οι πράκτορες διατηρούν ιστορικό αλληλεπιδράσεων με άλλους πράκτορες και με βάση την ικανότητα τους αυτή μπορούν να κρίνουν αν θα συνεργαστούν ή όχι.
- Η δυαδικότητα του αποτελέσματος μιας συνεργασίας. Δυο πράκτορες είτε θα συνεργαστούν είτε όχι.
- Ένας πράκτορας δεν έχει ποτέ απόλυτη γνώση για την αξιοπιστία ενός πράκτορα που θα συνεργαστεί. Αυτό πρακτικά σημαίνει ότι ακόμη και αν ένας πράκτορας είναι απόλυτα αξιόπιστος σε όλες τις παρελθοντικές συναλλαγές, στην επόμενη συναλλαγή η αξιοπιστία του δεν θεωρείται δεδομένη.

2.5.2 Κατηγορίες και αναπαράσταση εμπιστοσύνης

Στο σημείο αυτό παρουσιάζονται οι έννοιες και τα σύμβολα που χρησιμοποιούνται για την αναπαράσταση και τον υπολογισμό της εμπιστοσύνης.

Αναπαράσταση πρακτόρων

Οι πράκτορες συμβολίζονται με μικρούς λατινικούς χαρακτήρες $a, b, c \dots$. Όλοι οι πράκτορες που ανήκουν σε μια κοινότητα πρακτόρων αναπαρίστανται από το σύνολο A . Υποσύνολα του συνόλου A αναπαρίστανται ως $S_1, S_2 \dots$ τα οποία σύνολα ανήκουν στο σύνολο A . Τα υποσύνολα $S_1, S_2 \dots$ μιας κοινότητας πρακτόρων ορίζονται ως ένας αριθμός πρακτόρων που έχουν κάποιο κοινό χαρακτηριστικό – γνώρισμα.

Γνώση $K_x(y)$

Είναι πολύ βασικό να μπορούμε να αναπαραστήσουμε την πιθανότητα κάποιος πράκτορας να “γνωρίζει” κάποιον άλλον πράκτορα. Συγκεκριμένα, η γνώση αυτή συμβολίζεται ως $K_x(y)$.

$$K_x(y) = \begin{cases} 1 & , \text{ αν ο } x \text{ γνωρίζει τον } y \\ 0 & , \text{ αν ο } x \text{ δεν γνωρίζει τον } y \end{cases}$$

Βασική εμπιστοσύνη T_x

Υπάρχει ένα μέτρο που δείχνει κατά πόσο ένας πράκτορας είναι διατεθειμένος να εμπιστευτεί και κατόπιν να συναλλαχθεί με κάποιον άλλον πράκτορα. Αυτή η τάση ενός πράκτορα x να εμπιστεύεται τον οποιονδήποτε άλλο πράκτορα ονομάζεται βασική εμπιστοσύνη. Η βασική εμπιστοσύνη λαμβάνει τιμές στο διάστημα $[-1, 1]$. Η βασική εμπιστοσύνη επηρεάζεται από τις προηγούμενες εμπειρίες του πράκτορα με την έννοια ότι θετικά αποτελέσματα συναλλαγών (θετική εμπειρία) οδηγούν σε μεγαλύτερη προδιάθεση για εμπιστοσύνη και κατ’ επέκταση σε μεγαλύτερη τιμή της βασικής εμπιστοσύνης.

Γενική εμπιστοσύνη $T_x(y)$

Συμβολίζουμε με $T_x(y)$ την διάθεση του x να εμπιστευτεί τον y , όπου $-1 \leq T_x(y) < 1$. Η τιμή -1 δηλώνει ότι ο x δεν εμπιστεύεται καθόλου τον y (αρνητική εμπιστοσύνη), η τιμή 0 δηλώνει ότι ο x δεν έχει καθόλου γνώση για τον y (έλλειψη γνώσης) και η τιμή 1 συμβολίζει μια ιδεατή κατάσταση (απόλυτη εμπιστοσύνη) η οποία δεν ανταποκρίνεται στην πραγματικότητα και για το λόγο αυτό η $T_x(y)$ δεν λαμβάνει ποτέ τη τιμή αυτή. Η γενική εμπιστοσύνη δεν σχετίζεται με κάποια συγκεκριμένη κατάσταση, αλλά συμβολίζει τη γενικότερη εμπιστοσύνη του x προς τον y , δηλαδή την προσδοκία του x ότι ο y θα συμπεριφερθεί βάσει των συμφερόντων του x και δε θα τον βλάψει. Τέλος, αξίζει να σημειωθεί η όχι και τόσο ευδιάκριτη διαφορά μεταξύ της αρνητικής εμπιστοσύνης και της μηδενικής εμπιστοσύνης, αποτελώντας ένα από τα αρνητικά του μοντέλου.

Εμπιστοσύνη περίπτωσης $T_x(y, a)$

Οι πράκτορες λειτουργούν και δραστηριοποιούνται σε ένα διαρκώς μεταβαλλόμενο περιβάλλον. Αυτό σημαίνει ότι οι συνθήκες συναλλαγής των πρακτόρων συνεχώς αλλάζουν. Το αποτέλεσμα της συναλλαγής μεταξύ δύο πρακτόρων μπορεί να διαφέρει όταν αυτή λαμβάνει χώρα σε διαφορετικά περιβάλλοντα. Για παράδειγμα, μπορεί να εμπιστευόμαστε τον οποιοδήποτε ηλεκτρολόγο να μας επιδιορθώσει

μια μικρή ηλεκτρολογική βλάβη στο σπίτι, αλλά για μια μεγάλη ηλεκτρολογική εγκατάσταση θα επιλέγαμε έναν έμπειρο ηλεκτρολόγο με πολύ καλές συστάσεις. Οι τιμές που παίρνει η εμπιστοσύνη περίπτωσης ανήκουν στο διάστημα $[-1,1]$. Συμβολίζουμε με $T_x(y, a)$ την εμπιστοσύνη περίπτωσης όπου x, y οι πράκτορες και a η εκάστοτε περίπτωση στην οποία συναλλάσσονται.

Σπουδαιότητα $I_x(a)$

Σπουδαιότητα είναι η υποκειμενική εκτίμηση της έκβασης μιας συναλλαγής από τον πράκτορα x για την περίπτωση a . Με τον όρο υποκειμενική δηλώνεται ότι σε δύο διαφορετικές χρονικές στιγμές, ένας πράκτορας ίσως εκτιμήσει διαφορετικά την ίδια περίπτωση. Οι τιμές που λαμβάνει η σπουδαιότητα ανήκουν στο διάστημα $[0,1]$.

Χρησιμότητα $U_x(a)$

Η χρησιμότητα δείχνει το κατά πόσο ωφελημένος θα είναι ένας πράκτορας από την έκβαση μιας περίπτωσης a . Η χρησιμότητα λαμβάνει υπόψη όλες τις πιθανές εκβάσεις της περίπτωσης a , και είναι αυτό που τη χαρακτηρίζει αντικειμενική εν αντιθέσει με τη σπουδαιότητα. Η χρησιμότητα λαμβάνει τιμές στο πεδίο $[-1,1]$.

Όλα τα παραπάνω στοιχεία μπορούν συγκεντρωτικά να αναπαρασταθούν στον ακόλουθο πίνακα:

Περιγραφή	Αναπαράσταση	Πεδίο τιμών
Πράκτορες	a, b, c, \dots	
Σύνολο πρακτόρων	A	
Κοινωνίες πρακτόρων	$S_1, S_2, S_3, \dots \in A$	
Γνώση	$K_x(y)$	$[0,1]$
Σπουδαιότητα	$I_x(a)$	$[0,1]$
Χρησιμότητα	$U_x(a)$	$[-1,1]$
Βασική εμπιστοσύνη	T_x	$[-1,1]$
Γενική εμπιστοσύνη	$T_x(y)$	$[-1,1]$
Εμπιστοσύνη περίπτωσης	$T_x(y, a)$	$[-1,1]$

2.5.3 Υπολογισμός εμπιστοσύνης

Από τα τρία είδη εμπιστοσύνης (βασική, γενική, περίπτωσης) που εξετάσαμε, το μοντέλο του Marsh, πιο σημαντική θεωρεί την εμπιστοσύνη περίπτωσης $T_x(y, a)$. Αν η εμπιστοσύνη περίπτωσης ξεπερνά ένα όριο κατωφλίου, τότε η συναλλαγή μεταξύ των δύο πρακτόρων ολοκληρώνεται. Ορίζεται ως $\widehat{T_x(y, a)}$ η εκτίμηση της γενικής εμπιστοσύνης του πράκτορα x προς τον πράκτορα y για την περίπτωση a . Η εμπιστοσύνη περίπτωσης υπολογίζεται ως εξής :

$$T_x(y, a) = I_x(a)U_x(a)\widehat{T_x(y)} \quad (2.5.1)$$

Το αρνητικό στη σχέση αυτή είναι ότι η χρησιμότητα και η γενική εμπιστοσύνη μπορεί να λάβουν αρνητικές τιμές, οπότε το γινόμενο τους να δώσει θετικό αποτέλεσμα, πράγμα το οποίο δεν αντικατοπτρίζει την πραγματικότητα.

Για τον υπολογισμό της εκτίμησης της γενικής εμπιστοσύνης $\widehat{T_x(y)}$ υπάρχουν τριών ειδών προσεγγίσεις:

- *Βελτιστοποίηση εκτίμησης (αισιόδοξη προσέγγιση).* Με την προσέγγιση αυτή, ο πράκτορας x , για τη τιμή της γενικής εμπιστοσύνης λαμβάνει υπόψη τη μεγαλύτερη από τις προηγούμενες τιμές της γενικής εμπιστοσύνης $T_x(y)$.
- *Ελαχιστοποίηση εκτίμησης (απαισιόδοξη προσέγγιση).* Σε αντίθεση με παραπάνω, στην περίπτωση αυτή, λαμβάνεται υπόψη η μικρότερη τιμή της γενικής εμπιστοσύνης $T_x(y)$.
- *Ρεαλιστική προσέγγιση.* Κατά τη ρεαλιστική προσέγγιση η εκτίμηση της γενικής εμπιστοσύνης υπολογίζεται από τον μέσο όρο των προηγούμενων τιμών της γενικής εμπιστοσύνης. Δηλαδή:

$$\widehat{T_x(y)} = \frac{1}{|A|} \sum_{a \in A} T_x(y)$$

Όπου $|A|$ είναι το σύνολο των παρόμοιων καταστάσεων με αυτή της a , όπου ο πράκτορας x έχει αλληλεπιδράσει με τον y . Το σύνολο των προηγούμενων καταστάσεων που μπορούν να αποθηκευτούν μπορεί να περιορίζεται για δύο βασικούς λόγους. Πρώτον, να υπάρχει περιορισμός μνήμης στο σύστημα και δεύτερον, ο x να έχει πολύ λίγη ή καθόλου εμπειρία με τον y .

2.5.4 Κατώφλι συνεργασίας

Έχοντας υπολογίσει την εμπιστοσύνη περίπτωσης, πρέπει να οριστεί ένα μέτρο βάσει του οποίου θα κρίνεται αν ο x θα προχωρήσει σε συνεργασία με τον y ή όχι. Το μέτρο αυτό ονομάζεται κατώφλι συνεργασίας (cooperation threshold). Για να ολοκληρωθεί η συναλλαγή του x με τον y πρέπει :

$$T_x(y, a) > CooperationThreshold_x(a) \rightarrow Cooperate(x, y, a)$$

Ένας τρόπος υπολογισμού του κατωφλίου συνεργασίας είναι ο εξής:

$$CooperationThreshold_x(a) = \frac{PercievedRisk_x(a)}{PercievedCompetence_x(y, a)} I_x(a)$$

Όπου $PercievedRisk_x(a) \in [0, +\infty)$ είναι η αντίληψη του ρίσκου από τον x για την περίπτωση a και $PercievedCompetence_x(y, a) \in [0, +\infty)$ είναι η αντίληψη του x για την ικανότητα του y να ανταπεξέλθει στην περίπτωση a .

Το κατώφλι συνεργασίας είναι ανάλογο του ρίσκου και της σπουδαιότητας και αντιστρόφως ανάλογο της ικανότητας.

Για το ρίσκο ($PercievedRisk_x(a)$), ο Marsh δεν προτείνει κάποιο συγκεκριμένο τρόπο υπολογισμού. Ο υπολογισμός είναι δουλεία του πράκτορα x και ως εκ τούτου, το μέγεθος αυτό είναι υποκειμενικό. Συνεπώς ο υπολογισμός του κατωφλίου συνεργασίας πιθανόν να αποκλίνει από την πραγματικότητα, ειδικά στην περίπτωση που ο x δεν έχει ή έχει αλληλεπιδράσει πολύ λίγο κάτω από τις συνθήκες της περίπτωσης a .

Ο υπολογισμός της ικανότητας $PercievedCompetence_x(y, a)$ είναι πιο απλή υπόθεση. Υπάρχει η διάκριση σε τρεις κατηγορίες ανάλογα με το πόσο καλά γνωρίζει ο πράκτορας x τον πράκτορα y :

- Ο x έχει γνώση του y αλλά όχι σε παρόμοια περίπτωση a . Στην περίπτωση αυτή, χρησιμοποιείται η γνώση που έχει ο x για τον y από τις άλλες περιστάσεις β . Έτσι η ικανότητα υπολογίζεται ως εξής:

$$\begin{aligned} PercievedCompetence_x(y, a) &= \\ &= \frac{1}{|B|} \sum_{\beta \in B} ExperienceCompetence_x(y, \beta) \widehat{T_x(y)} \end{aligned}$$

Όπου $ExperienceCompetence_x(y, \beta)$ είναι η εμπειρία του x σχετικά με τον y για την περίπτωση β .

- Ο x έχει πλήρη άγνοια για τον y . Αυτή είναι σίγουρα η χειρότερη κατάσταση. Στην περίπτωση αυτή, μεγάλο ρόλο παίζει η γενική εμπιστοσύνη που έχει ο x στο να εμπιστεύεται τον οποιοδήποτε πράκτορα. Έτσι έχουμε:

$$PercievedCompetence_x(y, a) = T_x I_x(a)$$

- Ο x γνωρίζει καλά τον y και τον εμπιστεύεται. Στην περίπτωση αυτή ο x θα χρησιμοποιήσει την εμπειρία που έχει από τον y για την περίπτωση a . Συνεπώς :

$$\begin{aligned} PercievedCompetence_x(y, a) &= \\ &= \frac{1}{|A|} \sum_{a \in A} ExperienceCompetence_x(y, a) \widehat{T_x(y)} \end{aligned}$$

Έπειτα από τον υπολογισμό του κατωφλίου που πρότεινε αρχικά, ο Marsh παρατήρησε ότι το κατώφλι συνεργασίας θα πρέπει να είναι αντιστρόφως ανάλογο της $\widehat{T_x(y)}$. Έτσι, διαμόρφωσε την σχέση του κατωφλίου συνεργασίας ως εξής:

$$CooperationThreshold_x(a) = \frac{PercievedRisk_x(a)}{PercievedCompetence_x(y, a) + \widehat{T_x(y)}} I_x(a)$$

Από τη σχέση αυτή παρατηρούμε ότι εξαιτίας των αρνητικών τιμών που μπορεί να πάρει η εκτίμηση της γενικής εμπιστοσύνης $\widehat{T_x(y)}$, υπάρχει περίπτωση το άθροισμα του παρονομαστή να δώσει ποσότητα οριακά ή ίση με το μηδέν, πράγμα το οποίο φυσικά δεν είναι αποδεκτό.

Μια αναθεώρηση της παραπάνω σχέσης έχει να κάνει με το γεγονός ότι το όριο κατωφλίου μιας συναλλαγής πρέπει να είναι χαμηλό στην περίπτωση που η συναλλαγή αυτή έχει μεγάλη σπουδαιότητα για τον πράκτορα x. Δηλαδή, αλλάζει ριζικά η οπτική γωνία με την οποία αξιολογείται μια συναλλαγή, και για το λόγο αυτό η σπουδαιότητα μετακινείται στον παρονομαστή. Στην περίπτωση αυτή, το όριο κατωφλίου υπολογίζεται ως εξής:

$$CooperationThreshold_x(a) = \frac{PercievedRisk_x(a)}{(PercievedCompetence_x(y, a) + \widehat{T_x(y)}) I_x(a)}$$

2.5.5 Μνήμη

Όπως γίνεται αντιληπτό από τα παραπάνω, οι πράκτορες προκειμένου να αποθηκεύουν τα αποτελέσματα των αλληλεπιδράσεων τους με τους άλλους πράκτορες, πρέπει να έχουν επαρκή αποθηκευτικό χώρο. Φυσικά, όση περισσότερη μνήμη διαθέτει ένας πράκτορας, τόσο περισσότερα αξιόπιστα αποτελέσματα εξάγονται. Ο περιορισμός όμως της μνήμης είναι ένα μεγάλο μειονέκτημα.

2.5.6 Ενημέρωση εμπιστοσύνης

Ο Marsh υλοποιεί την ανανέωση της εμπιστοσύνης με έναν απλό τρόπο. Το ποσό αύξησης ή μείωσης της εμπιστοσύνης του x προς τον y εξαρτάται από την εκάστοτε περίσταση α.

Εάν ο y συνεργάστηκε με τον x στην περίπτωση α, τότε η γενική εμπιστοσύνη του x προς τον y αυξάνεται κατά B :

$$Cooperated(y)^a \rightarrow \begin{cases} T_x(y)^{a+1} > T_x(y)^a \\ T_x(y)^{a+1} = T_x(y)^a + B \end{cases}$$

Όπου B το όφελος για τον πράκτορα A από την περίπτωση α.

Και αντίστροφα, αν ο x δεν συνεργάστηκε με τον y ισχύει :

$$Defects(y)^a \rightarrow \begin{cases} T_x(y)^{a+1} < T_x(y)^a \\ T_x(y)^{a+1} = T_x(y)^a - C \end{cases}$$

Όπου C το κόστος για τον πράκτορα A από την κατάσταση α.

Εάν ληφθεί υπόψη η ανταποδοτικότητα του γ, κατά πόσο δηλαδή θα βοηθήσει τον x εφόσον ο x έχει βοηθήσει τον γ στο παρελθόν, τότε ισχύει η γενική εμπιστοσύνη αυξάνεται:

$$Helped(x, y, a)^{t-\delta} \wedge Cooperated(y, \beta)^t \rightarrow T_x(y)^{t+1} \geq T_x(y)^t$$

Στην αντίθετη περίπτωση, αν ο γ δεν συνεργαστεί με τον x, παρόλο που ο x στο παρελθόν είχε βοηθήσει τον γ, τότε η γενική εμπιστοσύνη μειώνεται κατά πολύ:

$$Helped(x, y, a)^{t-\delta} \wedge Defected(y, \beta)^t \rightarrow T_x(y)^{t+1} \ll T_x(y)^t$$

2.5.7 Συμπεράσματα

Ο Marsh ήταν από τους πρώτους που προσπάθησαν να αποτυπώσουν με μαθηματικές σχέσεις την κοινωνιολογική μέχρι τότε έννοια της εμπιστοσύνης. Πολλά από τα επόμενα μοντέλα που ακολούθησαν είχαν σαν ερευνητική αφετηρία το δικό του μοντέλο. Τα πλεονεκτήματα του μοντέλου μπορούν να θεωρηθούν τα παρακάτω:

- Αντιμετωπίζει τον κάθε πράκτορα και τη κάθε περίπτωση συνεργασίας ξεχωριστά, λαμβάνοντας υπόψη τα χαρακτηριστικά τους.
- Κάνει διάκριση ανάμεσα στην έλλειψη εμπιστοσύνης (αρνητικές τιμές) και στην αδυναμία αξιολόγησης (μηδενική τιμή).
- Ορίζει με μαθηματικά σύμβολα σχετικά πιο αυθαίρετες έννοιες όπως γνώση ($K_x(y)$), σημαντικότητα ($I_x(a)$), χρησιμότητα ($U_x(a)$) και άλλα, και με τον τρόπο αυτό καταφέρνει να τα ενσωματώσει σε μαθηματικές εξισώσεις για τον υπολογισμό της εμπιστοσύνης.

Ωστόσο διακρίνονται και μερικά αρνητικά σημεία στο μοντέλο αυτό:

- Όπως είδαμε στον υπολογισμό του ορίου κατωφλίου, ο παρονομαστής μπορεί να μηδενιστεί δίνοντας μη έγκυρο αποτέλεσμα. Κάτι παρόμοιο συμβαίνει και στον υπολογισμό της εμπιστοσύνης περίπτωσης όπου το γινόμενο της χρησιμότητας και της εκτίμησης της γενικής εμπιστοσύνης μπορεί να δώσει θετικό αποτέλεσμα, παρόλο που οι τιμές αυτών των δύο μεταβλητών να είναι αρνητικές.
- Υπάρχει σε μεγάλο βαθμό το στοιχείο της υποκειμενικότητας, καθώς ένας πράκτορας με δεδομένη συμπεριφορά μπορεί από έναν άλλο πράκτορα να θεωρείται αξιόπιστος και από κάποιον άλλον αναξιόπιστος.
- Ένα πολύ μεγάλο αρνητικό στοιχείο στο μοντέλο του Marsh είναι ότι δε λαμβάνει υπόψη τη φήμη που έχει σχηματιστεί για κάποιον πράκτορα από

τους άλλους πράκτορες που συνεργάστηκαν μαζί του. Έτσι οι εμπειρίες των άλλων πρακτόρων αν και είναι πολύ σημαντικές δεν αξιοποιούνται. Με τον τρόπο αυτό, κάθε πράκτορας λειτουργεί ατομικά μέσα στη κοινότητα πρακτόρων.

2.6 Το Μοντέλο των L.Xiong και L.Liu (PeerTrust)

Σε ένα Peer-to-Peer(P2P) σύστημα, υπάρχουν κοινότητες ομοίων (peer) πρακτόρων σε μια κοινωνία πρακτόρων. Συχνά, σε μια τέτοια κοινωνία, οι πράκτορες μιας κοινότητας αλληλεπιδρούν με τους πράκτορες μιας άλλης κοινότητας πρακτόρων. Στην εργασία τους, οι L.Xiong και L.Liu [14], προσπαθούν να εισάγουν έναν απλό τρόπο υπολογισμού της εμπιστοσύνης σε μια τέτοια κοινωνία πρακτόρων. Όπως ίσως γίνεται αντιληπτό, το μοντέλο αυτό ανήκει στις κατηγορίες των καταναμημένων συστημάτων και στα μοντέλα φήμης. Παρακάτω στην εργασία θα παρουσιαστούν οι τεχνικές και οι μηχανισμοί που χρησιμοποιεί το μοντέλο για τον υπολογισμό της εμπιστοσύνης.

2.6.1 Εισαγωγή σε ένα Peer-to-Peer σύστημα

Σε ένα ανοιχτό P2P σύστημα πρακτόρων, ένας πράκτορας μπορεί να εισέλθει και να εξέλθει οποιαδήποτε στιγμή. Συνεπώς, το να διατηρηθεί κάποια κεντρική γνώση για την αξιοπιστία ενός πράκτορα δεν είναι πάντα εφικτό. Έτσι, η έννοια της εμπιστοσύνης στα P2P συστήματα αποκτά ένα ιδιαίτερο νόημα, καθώς η εμπιστοσύνη για ένα συγκεκριμένο πράκτορα σε δυο διαφορετικές χρονικές στιγμές μπορεί να έχει τελείως διαφορετική τιμή.

Τα περισσότερα συστήματα εμπιστοσύνης εστιάζουν στο πώς να υπολογίσουν άμεσα την εμπιστοσύνη βάσει της προσωπικής τους εμπειρίας με τον πράκτορα αυτόν. Στα P2P συστήματα, λαμβάνονται υπόψη οι απόψεις όλων των πρακτόρων της κοινότητας στην οποία ανήκει ο πράκτορας που επιθυμεί συναλλαγή.

Υπάρχουν τρεις κύριοι παράγοντες που μπορούν να επηρεάσουν τον υπολογισμό της εμπιστοσύνης για κάποιο πράκτορα στο P2P μοντέλο:

- **Ικανοποίηση (satisfaction)**

Η ανατροφοδότηση που δίνει ένας πράκτορας σχετικά με το πόσο ικανοποιημένος έμεινε από την συνεργασία του με κάποιον άλλον πράκτορα είναι πολύ σημαντική για ένα P2P σύστημα. Η ικανοποίηση είναι ανάλογη της συμπεριφοράς του άλλου πράκτορα. Έτσι, μια μεγάλη ικανοποίηση είναι το αποτέλεσμα μιας καλής συνεργασίας και αντίστοιχα, μια χαμηλή ικανοποίηση είναι το αποτέλεσμα μιας κακής συνεργασίας.

- **Το πλήθος των συναλλαγών (number of interactions)**

Λαμβάνοντας υπόψη μόνο το πλήθος των συναλλαγών, μπορεί να προκύψουν εσφαλμένα αποτελέσματα όσον αφορά την αξιοπιστία ενός πράκτορα. Αυτό, μπορεί εύκολα να γίνει κατανοητό με ένα απλό παράδειγμα. Έστω σε ένα σύστημα ένας πράκτορας Α με 30 συναλλαγές να έχει 30 θετικές αξιολογήσεις και στο ίδιο σύστημα ένας άλλος πράκτορας με 100 συναλλαγές εκ των οποίων 30 θετικές και 70 αρνητικές αξιολογήσεις. Γίνεται εύκολα αντιληπτό ότι ο πράκτορας Α είναι πολύ πιο αξιόπιστος από τον πράκτορα Β. Ακόμη ένα παράδειγμα είναι ο πράκτορας Α έχει 100 αλληλεπιδράσεις με 50 αρνητικές και 50 θετικές αξιολογήσεις και ο πράκτορας Β με 50 αλληλεπιδράσεις και 25 θετικές και 25 αρνητικές αξιολογήσεις. Στην περίπτωση αυτή, αν και ο πράκτορας Α έχει πιο πολλές θετικές αξιολογήσεις από τον πράκτορα Β, η αναλογία θετικών προς αρνητικών είναι ίδια και στις δύο περιπτώσεις. Τέτοιου είδους παρανοήσεις συμβαίνουν όταν λαμβάνουμε υπόψη μόνο το πλήθος των θετικών αξιολογήσεων.

- **Εξισορροπητικός παράγοντας εμπιστοσύνης (balance trust factor)**

Η αξιολόγηση στην οποία θα ανατρέξει κάποιο μέλος της κοινωνίας ίσως είναι κακόβουλη ή παραπλανητική. Μπορεί για παράδειγμα για χαμηλής αξίας συναλλαγές ένας πράκτορας να συμπεριφέρεται πολύ καλά και αντίστοιχα στις υψηλής αξίας συναλλαγές να συμπεριφέρεται αναξιόπιστα. Στην περίπτωση που οι χαμηλής αξίας συναλλαγές είναι σημαντικά περισσότερες από τις σημαντικές αλλαγές, ο πράκτορας αυτός θα έχει χτίσει το προφίλ ενός καλού και αξιόπιστου πράκτορα. Για την περίπτωση τέτοιων δυσλειτουργιών, το μοντέλο προτείνει την εισαγωγή ενός μηχανισμού που σε κάθε συναλλαγή θα ενσωματώνει τα ιδιαίτερα χαρακτηριστικά κάθε συναλλαγής όπως τη σπουδαιότητα, την επικινδυνότητα και αλλά ανάλογα με τη φύση της κάθε συναλλαγής.

2.6.2 Υπολογισμός εμπιστοσύνης

Στο σημείο αυτό θα παρουσιάσουμε τον τρόπο υπολογισμού της εμπιστοσύνης για το μοντέλο Peer Trust.

Έστω P το σύνολο N πρακτόρων και u, v είναι δύο πράκτορες του συνόλου P . Όπως έχουμε ήδη αναφέρει το σύστημα μας είναι ανοιχτό, δηλαδή ο συνολικός αριθμός των πρακτόρων που ανήκουν στο σύνολο P διαρκώς μεταβάλλεται. Έστω $I(u, v, t)$ είναι το σύνολο των αλληλεπιδράσεων που είχε ο πράκτορας u με τον πράκτορα v μέχρι την χρονική στιγμή t . Έστω $S(u, v, t)$ το πόσο ικανοποιημένος είναι ο πράκτορας u από την αλληλεπίδραση του με τον πράκτορα v μέχρι την t χρονική στιγμή. Έστω $C_r(v, t)$ ο παράγοντας σταθεροποίησης της εμπιστοσύνης (balance trust factor) που αντισταθμίζει τον κίνδυνο των μη αξιόπιστων ανατροφοδοτήσεων.

Το μοντέλο για την αξιοπιστία του πράκτορα u προς τους άλλους πράκτορες προτείνει την σχέση:

$$T(u, t) = \frac{\sum_{v \in P, u \neq v} S(u, v, t) C_r(v, t)}{\sum_{v \in P, u \neq v} I(u, v, t)}$$

Με $T(u, t) \in [0, 1]$

Από τη σχέση αυτή συμπεραίνουμε ότι όσο μεγαλύτερη είναι η τιμή της $T(u, t)$, τόσο πιο αξιόπιστος είναι ο πράκτορας u . Πρέπει να σημειωθεί ότι η τιμή της $T(u, t)$ δεν δείχνει το ποσό της εμπιστοσύνης κάποιου συγκεκριμένου πράκτορα για τον πράκτορα u , αλλά είναι ένα μέτρο που βοηθάει τους πράκτορες να σχηματίσουν μια εικόνα για την αξιοπιστία του u . Αξίζει να σημειωθεί επίσης, ότι διαφορετικοί πράκτορες ίσως έχουν διαφορετική αντίληψη για την ίδια τιμή της εμπιστοσύνης $T(u, t)$ και αυτό οφείλεται στο ότι κάθε πράκτορας έχει διαφορετικό σύστημα αξιολόγησης. Στην ουσία, μπορούμε να θεωρήσουμε ότι η τιμή $T(u, t)$ είναι η φήμη που δημιουργεί ο πράκτορας u στους άλλους πράκτορες.

Αν θεωρήσουμε ότι $C(u, v, t)$ είναι ο αριθμός των παραπόνων που ο πράκτορας u λαμβάνει από τον πράκτορα v μέχρι τη χρονική στιγμή t , τότε ισχύει ότι :

$$C(u, v, t) = I(u, v, t) - S(u, v, t) \quad (2)$$

Ας εξετάσουμε τον παράγοντα σταθεροποίησης της εμπιστοσύνης που είδαμε παραπάνω. Είναι προφανές ότι οι πράκτορες δε μπορούν να έχουν τυφλή εμπιστοσύνη για τις ανατροφοδοτήσεις που λαμβάνουν όσον αφορά την αξιολόγηση πρακτόρων από άλλους πράκτορες. Γενικά, είναι πιθανό, ένας όχι και τόσο αξιόπιστος πράκτορας να αξιολογήσει κάποιον άλλο πράκτορα με μια κακή κριτική (ψευδή παράπονα) για να κρύψει τη δική του άσχημη συμπεριφορά. Για να διαχειριστούμε μια τέτοια ενέργεια, γίνεται μια απλή υπόθεση: όλοι οι αξιόπιστοι πράκτορες δεν κάνουν ψευδή παράπονα και οι αναξιόπιστοι κάνουν ψευδή παράπονα (όταν δε συμπεριφέρονται σωστά σε μία συναλλαγή). Όποτε, σύμφωνα με την παραπάνω παραδοχή, τα παράπονα από έναν αξιόπιστο πράκτορα λαμβάνονται υπόψη και τα παράπονα από έναν αναξιόπιστο δε λαμβάνονται υπόψη.

Το γινόμενο $S(u, v, t) \cdot C_r(v, t)$ δηλώνει την εξισορροπημένη ποσότητα της ικανοποίησης από τον πράκτορα v για τον u . Το γινόμενο $C(u, v, t) \cdot T(v, t)$ δείχνει το ποσό αξιοπιστίας είναι ο v στις αξιολογήσεις του μέχρι την t χρονική στιγμή.

Η σχέση (2) μπορεί να γραφτεί και ως:

$$S(u, v, t) = I(u, v, t) - C(u, v, t) \quad (3)$$

Αν στη σχέση τροποποιηθεί και αντί για την ικανοποίηση $S(u, v, t)$ τοποθετήσουμε την ποσότητα $S(u, v, t) \cdot C_r(v, t)$ και τον αριθμό παραπόνων $C(u, v, t)$ τον αντικαταστήσουμε με την ποσότητα $C(u, v, t) \cdot T(v, t)$, τότε η σχέση (3) γράφεται ως εξής:

$$S(u, v, t) \cdot C_r(v, t) = I(u, v, t) - C(u, v, t) \cdot T(v, t) \quad (4)$$

Με αντικατάσταση της σχέσης (4) στη σχέση (1), η (1) γίνεται:

$$\begin{aligned} T(u, t) &= \frac{\sum_{v \in P, u \neq v} (I(u, v, t) - C(u, v, t) \cdot T(v, t))}{\sum_{v \in P, u \neq v} I(u, v, t)} \\ &= \frac{\sum_{v \in P, u \neq v} I(u, v, t)}{\sum_{v \in P, u \neq v} I(u, v, t)} - \frac{\sum_{v \in P, u \neq v} C(u, v, t) \cdot T(v, t)}{\sum_{v \in P, u \neq v} I(u, v, t)} \\ &= 1 - \frac{\sum_{v \in P, u \neq v} C(u, v, t) \cdot T(v, t)}{\sum_{v \in P, u \neq v} I(u, v, t)} \in [0, 1] \quad (5) \end{aligned}$$

Η σχέση αυτή μας δίνει ότι και η σχέση (1) με την διαφορά ότι είναι τροποποιημένη και βασισμένη στα παράπονα $C(u, v, t)$ των πρακτόρων.

Το μοντέλο PeerTrust στη συνέχεια διακρίνει δύο υπολογιστικές μεθόδους για τον υπολογισμό της εμπιστοσύνης . Πρώτη μέθοδος είναι ο δυναμικός υπολογισμός , όπου χρησιμοποιεί τα πιο φρέσκα δεδομένα για τον υπολογισμό της εμπιστοσύνης. Η δεύτερη μέθοδος είναι ο προσεγγιστικός υπολογισμός , όπου χρησιμοποιεί την έννοια της cache για να επιταχύνει τη διαδικασία υπολογισμού της εμπιστοσύνης. Παρακάτω θα εξετάσουμε τις δύο αυτές υπολογιστικές μεθόδους.

Δυναμικός υπολογισμός (dynamic computation)

Τα μεγέθη $I(u, v, t)$ και $C(u, v, t)$ έχουν οριστεί ήδη. Έστω $I(u, t)$ είναι ο συνολικός αριθμός των αλληλεπιδράσεων του u με όλους τους άλλους πράκτορες. Έστω ότι $A = (a_{uv})$ είναι ένας τετραγωνικός πίνακας $N \times N$ (όπου N ο αριθμός πρακτόρων) με κάθε στοιχείο του πίνακα να υπολογίζεται ως εξής:

$$\alpha_{uv} = \begin{cases} \frac{C(u, v, t)}{I(u, t)} & , \quad I(u, t) \neq 0 \\ 0 & , \quad I(u, t) = 0 \end{cases}$$

Δηλαδή, κάθε στοιχείο a_{uv} του πίνακα A είναι ο λόγος των παραπόνων που έχει λάβει ο πράκτορας u από τον πράκτορα v προς τον συνολικό αριθμό των αλληλεπιδράσεων του πράκτορα u με όλους πράκτορες.

Έστω ότι $T = (t_u)$ είναι ένα διάνυσμα-στήλη του πίνακα A που περιέχει τις τιμές της εμπιστοσύνης όλων των πρακτόρων για τον πράκτορα u ως τη χρονική στιγμή t . Αν θέσουμε $t_u = T = T(u, t)$ και δεδομένης της παραπάνω σχέσης , η σχέση (5) γράφεται ως εξής:

$$T = 1 - AT$$

Παρατηρούμε το υψηλό υπολογιστικό κόστος της εμπιστοσύνης, πράγμα που κάνει τον υπολογισμό αυτό πολύπλοκο και χρονοβόρο. Για το λόγο αυτό, το μοντέλο εισάγει τον προσεγγιστικό υπολογισμό χρησιμοποιώντας την cache του κάθε πράκτορα.

Προσεγγιστικός υπολογισμός (approximate computation)

Με τη μέθοδο αυτή, ο κάθε πράκτορας διατηρεί μια cache όπου κρατάει όλες τις τιμές της εμπιστοσύνης που έχει από όλες τις αλληλεπιδράσεις με τους άλλους πράκτορες. Για το μέγεθος της cache το μοντέλο δε κάνει κάποια ιδιαίτερη αναφορά παρά μόνο αναφέρει ότι εξαρτάται από τους διαθέσιμους πόρους κάθε πράκτορα.

Όταν ο πράκτορας w πρέπει να υπολογίσει την εμπιστοσύνη για κάποιο άλλο πράκτορα u , ανατρέχει στις cache των άλλων πρακτόρων, αν βρει τιμές για τον u τις χρησιμοποιεί, αν όχι χρησιμοποιεί τις προκαθορισμένες τιμές. Όταν ο w υπολογίσει τη τιμή της εμπιστοσύνης για τον u , τοποθετεί τη τιμή αυτή στη δική του cache. Στην περίπτωση που η cache είναι γεμάτη, χρησιμοποιείται η πολιτική αντικατάστασης LRU (least recently used) για τη τοποθέτηση του νέου στοιχείου στη cache. Η ιδέα δηλαδή είναι πως χρησιμοποιείται ο πράκτορας w σαν ενδιάμεσος πράκτορας, του οποίου στην ουσία εκμεταλλευόμαστε την cache του.

Έστω $T_{cache}(v, t')$ η τιμή της αξιοπιστίας του v μέχρι τη t' χρονική στιγμή, η οποία βρίσκεται στη cache του w . Η προκαθορισμένη τιμή είναι $T_{default}$.

$$T(u, t) = 1 - \frac{\sum_{v \in P, u \neq v} C(u, v, t) \cdot T'(v, t)}{\sum_{v \in P, u \neq v} I(u, v, t)}$$

$$T'(v, t) = \begin{cases} T_{cache}(v, t') & , \text{ για cache hit} \\ T_{default} & , \text{ για cache miss} \end{cases}$$

2.6.3 Συμπεράσματα

Το μοντέλο PeerTrust συνδυάζει πολλούς παράγοντες εμπιστοσύνης και αυτό το καθιστά πολύ χρήσιμο εργαλείο στα χέρια του ερευνητή. Μερικά θετικά σημεία που χρίζουν αναφοράς τα οποία προέκυψαν από τη μελέτη του μοντέλου είναι τα εξής:

- Λαμβάνει υπόψη τα στοιχεία από όλους τους πράκτορες. Αυτό, επιτρέπει να διαμορφώνεται μια πιο σφαιρική και έγκυρη εικόνα για τον πράκτορα που αξιολογείται.
- Εισάγει και χρησιμοποιεί την έννοια των παραπόνων ($C(u, v, t)$). Είναι πολύ σημαντικό να μπορούν οι πράκτορες να εκφράσουν τα παράπονα τους για

κάποιο άλλο πράκτορα και το μοντέλο αυτό εισάγει την έννοια αυτή και την εισάγει στον υπολογισμό της εμπιστοσύνης.

- Με τον παράγοντα σταθεροποίησης της εμπιστοσύνης ($C_r(v, t)$) οι μη αξιόπιστες ανατροφοδοτήσεις περιορίζονται, καθιστώντας έτσι το μοντέλο πιο ακριβές.

Στα αρνητικά σημεία του μοντέλου μπορούμε να συμπεριλάβουμε τα εξής:

- Στον τύπο υπολογισμού της εμπιστοσύνης δε λαμβάνεται υπόψη η περίπτωση της ανυπαρξίας (έλλειψης) εμπιστοσύνης.
- Ο υπολογισμός της εμπιστοσύνης είναι αρκετά χρονοβόρος, καθώς υπάρχει μεγάλο κόστος λόγω της επικοινωνίας με όλους τους άλλους πράκτορες.
- Η χρήση προκαθορισμένων τιμών για την τιμή της εμπιστοσύνης στην περίπτωση που δεν έχουν αλληλεπιδράσει δύο πράκτορες μεταξύ τους, είναι δυνατό να οδηγήσουν σε συμπεράσματα που απέχουν πολύ από την πραγματικότητα.

Συμπερασματικά, το μοντέλο είναι αρκετά αξιόπιστο καθώς χρησιμοποιεί τις τιμές εμπιστοσύνης από όλους τους πράκτορες, αλλά εξαιτίας του μεγάλου κόστους επικοινωνίας μεταξύ των πρακτόρων, το μοντέλο είναι μη αποδοτικό από θέμα χρόνου.

2.7 Το μοντέλο των Mui-Mohtashemi-Halberstadt

Το μοντέλο των Mui, Mohtashemi και Halberstadt[15] αναπτύχθηκε το 2002 και στηρίζεται στην έννοια της αμοιβαιότητας. Με τον όρο αμοιβαιότητα στην κοινωνία των πρακτόρων εννοούμε την τάση που έχουν οι πράκτορες να συμπεριφέρονται σε κάποιον άλλον πράκτορα ανάλογα με το πώς τους έχει συμπεριφερθεί ο πράκτορας αυτός στο παρελθόν. Δηλαδή, όταν ένας πράκτορας είναι συνεργάσιμος αναμένει θετική συμπεριφορά, και όταν δεν είναι αναμένει κακή συμπεριφορά. Αξίζει να σημειωθεί ότι είναι το πρώτο μοντέλο που δίνει τόσο μεγάλη βάση στην αμοιβαιότητα.

2.7.1 Ορισμοί εννοιών

Οι Mui, Mohtashemi και Halberstadt ορίζουν στο μοντέλο τους βασικούς άξονες πάνω στους οποίους στηρίζονται για τον υπολογισμό της εμπιστοσύνης. Για το υπόλοιπο του μοντέλου θεωρούμε ότι το σύστημα στο οποίο το μοντέλο αναπτύσσεται είναι κλειστό (κανένας πράκτορας δεν εισέρχεται και δεν εξέρχεται από το σύστημα).

Αμοιβαιότητα (Reciprocity) $\gamma \in [0, 1]$

Με τον όρο αμοιβαιότητα εννοείται η διαδικασία αυτή κατά την οποία η δράση ενός πράκτορα Α σε μια συναλλαγή με κάποιον άλλο πράκτορα Β να επηρεάζεται σε ένα πολύ μεγάλο βαθμό από την συμπεριφορά του πράκτορα Β σε τυχόν παρελθοντική συνάντηση με τον πράκτορα Α. Με απλά λόγια, το αν θα συνεργαστεί ο πράκτορας Α με τον πράκτορα Β εξαρτάται από το αν ο πράκτορας Β στο παρελθόν είχε επιδείξει θετική συμπεριφορά.

Στη συνέχεια, το μοντέλο διακρίνει την αμοιβαιότητα σε δύο κατηγορίες: την άμεση και την έμμεση αμοιβαιότητα. Με τον όρο άμεση αμοιβαιότητα (direct reciprocity) εννοείται η “απευθείας” αμοιβαιότητα που αναπτύσσεται μεταξύ δύο πρακτόρων. Με τον όρο έμμεση αμοιβαιότητα (indirect reciprocity) εννοείται η αμοιβαιότητα που αναπτύσσεται μεταξύ δύο πρακτόρων όταν διαμεσολαβεί ένας ή περισσότεροι πράκτορες.

Την αμοιβαιότητα μπορούμε να την προσεγγίσουμε με δύο τρόπους. Αρχικά, μπορεί να θεωρηθεί σαν ένα κοινωνικό μέτρο, του οποίου όσο πιο μεγάλη είναι η τιμή του, τόσο πιο πιθανό είναι ο πράκτορας να δρα με τρόπο ανταποδοτικό. Έπειτα, μπορεί να θεωρηθεί σαν δυαδική αμοιβαιότητα μεταξύ δύο πρακτόρων. Όσο πιο μεγάλη είναι η τιμή της δυαδικής αμοιβαιότητας, τόσο πιο μεγάλη η προσδοκία να συμπεριφερθούν οι πράκτορες με τρόπο ανταποδοτικό στις μεταξύ τους συναλλαγές.

Με το σύμβολο γ αναπαριστούμε το πόσες αμοιβαίες πράξεις συμβαίνουν σε μια κοινωνία πρακτόρων. Δηλαδή, αναπαριστά το πλήθος των ενεργειών συνεργασίας που απαντώνται με πράξεις συνεργασίας, και των ενεργειών μη συνεργασίας που απαντώνται με πράξεις μη συνεργασίας.

Φήμη (Reputation) $\theta_{ij}(c) \in [0, 1]$

Είναι η αντίληψη που δημιουργείται από τις παρελθοντικές ενέργειες ενός πράκτορα σχετικά με την αξιοπιστία του. Ένας πράκτορας με καλή φήμη είναι πιο πιθανό να συνεργαστεί απ’ ότι ένας με κακή φήμη.

Με το σύμβολο $\theta_{ij}(c)$ αναπαριστούμε την πιθανότητα ο πράκτορας α_j να ανταποδώσει τις ενέργειες του πράκτορα α_i . Όσο μεγαλύτερη είναι η τιμή, τόσο μεγαλύτερη είναι η τάση για ανταπόδοση. Όπου c δηλώνεται το πλαίσιο που οι δύο πράκτορες αλληλεπιδρούν, το οποίο λαμβάνει τιμές από το σύνολο C που περιλαμβάνει όλες τις δυνατές περιπτώσεις μέσα στις οποίες δύναται να αλληλεπιδράσουν οι δύο πράκτορες μεταξύ τους.

Ιστορικό $D_{ij}(c)$

Με το σύμβολο αυτό αναπαριστούμε το ιστορικό των συναντήσεων μεταξύ δύο πρακτόρων στο πλαίσιο δράσης c .

Εμπιστοσύνη $\tau_{ij}(c) \in [0, 1]$

Τον όρο εμπιστοσύνη το μοντέλο τον ορίζει ως την υποκειμενική προσδοκία που έχει ένας πράκτορας για την μελλοντική συμπεριφορά ενός άλλου πράκτορα βάσει του ιστορικού των συναντήσεών τους.

Η εμπιστοσύνη εκφράζεται ως

$$\tau_{ij}(c) = E[\theta_{ij}(c)|D_{ij}(c)]$$

Δηλαδή, είναι η προσδοκία (expectation) της φήμης ενός πράκτορα δεδομένου του ιστορικού των συναντήσεων μεταξύ των δύο πρακτόρων.

2.7.2 Υπολογισμός εμπιστοσύνης

Έστω a και b οι πράκτορες που συνεργάζονται σε ένα πλαίσιο c . Από δω και στο εξής στο κεφάλαιο αυτό θεωρούμε ότι ο πράκτορας a επιθυμεί να εκτιμήσει την αξιοπιστία του πράκτορα b . Όπου θ_{ab} η φήμη του b στα μάτια του a . Η δυαδική μεταβλητή $x_{ab}(i)$ αναπαριστά την i -οστή συναλλαγή μεταξύ του a και του b .

$$x_{ab}(i) = \begin{cases} 1 & , \text{αν ο πράκτορας } b \text{ συνεργάστηκε} \\ 0 & , \text{αν ο πράκτορας } b \text{ δεν συνεργάστηκε} \end{cases}$$

Το σύνολο των τελευταίων n συναλλαγών μεταξύ του a και του b αναπαρίσταται ως εξής:

$$D_{ab} = \{x_{ab}(1), x_{ab}(2), \dots, x_{ab}(n)\}$$

Έστω p ο αριθμός των επιτυχών συνεργασιών του πράκτορα b με τον πράκτορα a σε n πλήθος συναντήσεων τους. Η φήμη θ_{ab} είναι μια συνάρτηση αναλογίας των μεταβλητών n και p . Σύμφωνα με την θεωρία της στατιστικής μια τυχαία μεταβλητή αναλογίας (proportion random variable) μπορεί να μοντελοποιηθεί σαν μια Β' κατανομή [16].

$$p(\hat{\theta}) = \text{Beta}(c_1, c_2)$$

Όπου $\hat{\theta}$ η εκτίμηση της συνάρτησης θ και c_1, c_2 είναι παράμετροι που καθορίζονται από προηγούμενες συναντήσεις. Ένας απλός εκτιμητής $\hat{\theta}$ για την τυχαία μεταβλητή θ είναι ο εξής:

$$\hat{\theta}_{ab} = \frac{p}{n}$$

Δεδομένου ότι η πιθανότητα ο b να συνεργαστεί με τον α είναι ανεξάρτητη από το αποτέλεσμα των προηγούμενων συναντήσεων τους, τότε η πιθανότητα να έχουμε p συνεργασίες σε σύνολο n συναντήσεων (άρα n-p αποτυχίες συνεργασίας) μπορεί να μοντελοποιηθεί ως εξής:

$$L(D_{ab}|\hat{\theta}) = \hat{\theta}^p (1 - \hat{\theta})^{n-p}$$

Η B' κατανομή γίνεται ο πρότερος συζυγής (conjugate prior) αυτής της πιθανότητας [17], έτσι η επόμενη εκτίμηση για το $\hat{\theta}$ γίνεται:

$$P(\hat{\theta}|D) = \text{Beta}(c_1 + p, c_2 + n - p)$$

Από την μέση τιμή της B' κατανομής έχουμε ότι:

$$E(\hat{\theta}|D) = \frac{c_1 + p}{c_1 + p + c_2 + n - p} = \frac{c_1 + p}{c_1 + c_2 + n}$$

Για την διακύμανση της B' κατανομής έχουμε ότι:

$$\sigma_{\hat{\theta}|D}^2 = \frac{(c_1 + p)(c_2 + n - p)}{(c_1 + c_2 + n - 1)(c_1 + c_2 + n)^2}$$

Στην επόμενη συνάντηση η εκτίμηση του α ότι ο b θα συνεργαστεί είναι:

$$\tau_{ab} = p(x_{ab}(n+1) = 1|D_{ab}) = E[\widehat{\theta}_{ab}|D_{ab}]$$

Σύμφωνα με το θεώρημα του ορίου του Chernoff[18] ισχύει ότι, αν $X_{ab}(1), X_{ab}(2), \dots, X_{ab}(m)$ είναι μια ακολουθία από m ανεξάρτητες δοκιμές Bernoulli, κάθε μια με πιθανότητα επιτυχίας $E[X_{ab}] = \theta$ και ορίσουμε τον ακόλουθο εκτιμητή $\hat{\theta} = (X_{ab}(1) + X_{ab}(2) + \dots + X_{ab}(m))/m$, όπου θ είναι μια τυχαία μεταβλητή που αντιπροσωπεύει το ποσοστό της επιτυχίας έτσι ώστε $E[\hat{\theta}] = \theta$, τότε για $0 \leq \varepsilon \leq 1$ και $0 \leq \delta \leq 1$ ισχύει το όριο :

$$\Pr[|\theta - \hat{\theta}| \geq \varepsilon] \leq 2e^{-2m\varepsilon^2} \leq \delta$$

Όπου ε στην ουσία είναι η απόκλιση της πραγματικής παραμέτρου θ από τον εκτιμητή $\hat{\theta}$, το m αναπαριστά τον ελάχιστο αριθμό αλληλεπιδράσεων μεταξύ του α και του b, έτσι ώστε να υπάρξει το ελάχιστο επίπεδο εσφαλμένης εκτίμησης. Το m υπολογίζεται ως εξής:

$$m \geq -\frac{1}{2\varepsilon^2} \ln\left(\frac{\delta}{2}\right)$$

Έστω $\gamma_c = 1 - \delta$. Το γ_c εκφράζει ένα μέτρο ασφάλειας του εκτιμητή $\hat{\theta}$. Όσο το γ_c τείνει στο 1, τόσο πιο ακριβής είναι η εκτίμηση της εμπιστοσύνης. Το ότι το γ_c τείνει στο 1 σημαίνει επίσης ότι το m είναι ικανώς μεγάλο ώστε να επιτευχθεί το επιθυμητό επίπεδο λάθους ϵ . Το γ_c μπορεί να επιλεγεί από τον χρήστη του μοντέλου και στην ουσία αναφέρεται στην αξιοπιστία της φήμης του πράκτορα b .

Όπως είδαμε και παραπάνω, το μοντέλο διακρίνει δύο ειδών αμοιβαιότητες, τη δυαδική και την κοινωνική. Στο σημείο αυτό οι δύο αυτές έννοιες υπολογίζονται με μαθηματική έκφραση.

$$dyadicreciprocity(a, b) = \frac{mutualcooperation(a, b) + mutualdefection(a, b)}{all\ encounters(a, b)}$$

$$socialreciprocity = \frac{allmutualcooperation + allmutualdefection}{all\ encounters\ in\ society}$$

Αν γ_{ab} είναι η δυαδική αμοιβαιότητα μεταξύ των πρακτόρων a και b , τότε αν $\gamma_{ab} \leq \gamma_c$ η φήμη και η εμπιστοσύνη δεν θεωρούνται αξιόπιστες.

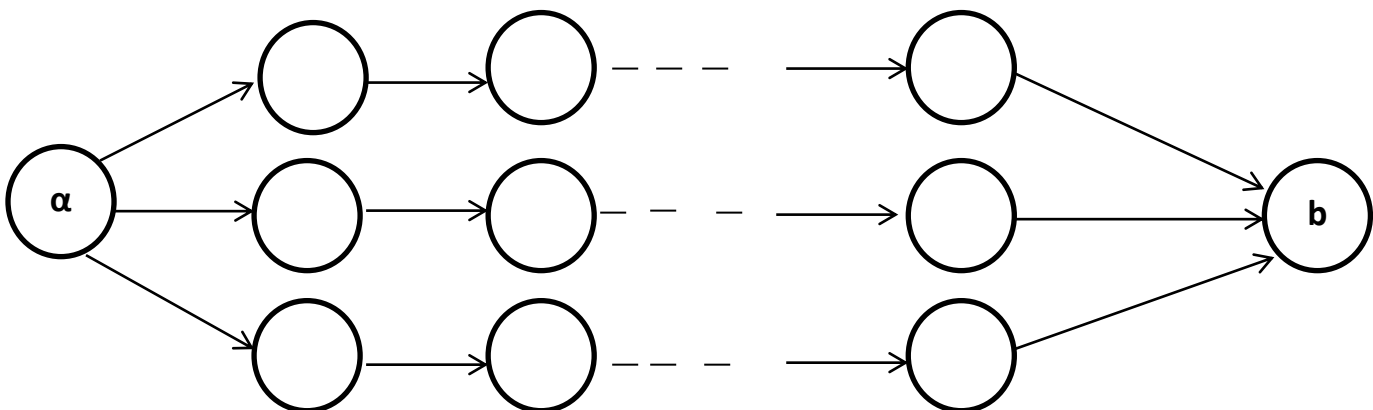
Στη περίπτωση που οι πράκτορες a και b είναι τελείως άγνωστοι μεταξύ τους, δηλαδή να μην έχουν αλληλεπιδράσει ποτέ και να μην έχουν κοινούς γνωστούς, τότε η εκτίμηση της φήμης θεωρείται μια μεταβλητή ομοιόμορφα κατανεμημένη

$$p(\hat{\theta}) = \begin{cases} 1, & 0 < \hat{\theta} < 1 \\ 0, & \text{αλλιώς} \end{cases}$$

2.7.3 Η διάδοση της φήμης

Στο κεφάλαιο αυτό μελετάται ο τρόπος με τον οποίο η φήμη διαδίδεται σε μια κοινωνία πρακτόρων με αλυσίδες πεπερασμένου αριθμού πρακτόρων.

Το παρακάτω σχήμα αναπαριστά τον τρόπο διάδοσης της φήμης για το μοντέλο.



Όπως παρατηρούμε, το κύριο χαρακτηριστικό του διαγράμματος είναι ότι οι αλυσίδες είναι παράλληλες μεταξύ τους.

Έστω ότι ο πράκτορας α επιθυμεί να υπολογίσει την αξιοπιστία του πράκτορα b μέσω της αλυσίδας των πρακτόρων που παρεμβάλλονται μεταξύ τους. Έστω ότι α λαμβάνει πληροφορίες για την αξιοπιστία του b μέσα από k διαφορετικές διαδρομές. Αρχικά θα υπολογίσουμε την αξιοπιστία για κάθε μία ξεχωριστή αλυσίδα. Έπειτα κάθε αλυσίδα θα συνεισφέρει στον τελικό υπολογισμό της εμπιστοσύνης με την ανάλογη βαρύτητα.

Το βάρος για μία σύνδεση μεταξύ δύο ενδιαμέσων πρακτόρων i και j είναι

$$w_{ij} = \begin{cases} \frac{m_{ij}}{m}, m_{ij} < m \\ 1, \text{διαφορετικά} \end{cases}$$

Όπου m_{ij} το πλήθος των συναντήσεων μεταξύ του i και του j και m ο ελάχιστος αριθμός αλληλεπιδράσεων μεταξύ i και j έτσι ώστε η εκτίμηση της εμπιστοσύνης να θεωρείται αξιόπιστη.

Το συνολικό βάρος μιας αλυσίδας είναι το γινόμενο των βαρών των συνδέσεων που απαρτίζουν την αλυσίδα.

$$w_i = \prod_{j=1}^{l_i} w_{ij}$$

Όπου l_i το μήκος του μονοπατιού i και w_{ij} το βάρος του j -οστού τμήματος της i -όστης αλυσίδας.

Έχοντας υπολογίσει όλα τα βάρη για όλες τις αλυσίδες μεταξύ των πρακτόρων α και b , η συνολική φήμη r_{ab} που έχει ο πράκτορας α για τον b , υπολογίζεται ως εξής:

$$r_{ab} = \sum_{i=1}^n r_{ab}(i) \overline{w}_i$$

Όπου $r_{ab}(i)$ η φήμη του επιμέρους μονοπατιού σε κανονικοποιημένη μορφή (θα ισχύει ότι $\sum_{i=1}^n \overline{w}_i = 1$) και \overline{w}_i το κανονικοποιημένο βάρος του μονοπατιού i .

2.7.4 Συμπεράσματα

Οι Mui, Mohtashemi και Halberstadt χρησιμοποίησαν εκτενώς τη θεωρία των πιθανοτήτων για να υπολογίσουν την εμπιστοσύνη σε μια κοινωνία πρακτόρων. Πολύ μεγάλη σημασία δόθηκε στην αμοιβαιότητα και πως αυτή επηρεάζει τις μελλοντικές κινήσεις των πρακτόρων.

Στα θετικά του μοντέλου μπορούν να συμπεριληφθούν τα κάτωθι:

- Διακρίνει την έννοια της εμπιστοσύνης από αυτή της φήμης.
- Το μοντέλο είναι χρηστικό καθώς δεν περιλαμβάνει πολύπλοκους υπολογισμούς.
- Εισάγει στον υπολογισμό της εμπιστοσύνης έννοιες που τα περισσότερα μοντέλα δεν έχουν συμπεριλάβει όπως αυτή της αμοιβαιότητας, διαχωρίζοντάς την έπειτα σε δυαδική και κοινωνική.

Στα αρνητικά σημεία του μοντέλου μπορούμε να διακρίνουμε τα εξής:

- Κάνει την παραδοχή ότι όλα τα μονοπάτια διάδοσης της φήμης είναι παράλληλα, πράγμα το οποίο στην πραγματικότητα ίσως είναι εφικτό μόνο σε μεγάλης κλίμακας κοινωνίες πρακτόρων, χωρίς οι πράκτορες να έχουν πολλές αλληλεπιδράσεις με άλλους πράκτορες.
- Υπάρχουν πολλά ασαφή σημεία, όπως για παράδειγμα δεν αναφέρεται πώς βρίσκουμε τα μονοπάτια διάδοσης της φήμης
- Υπάρχουν μόνο δύο δυνατά αποτελέσματα σε μια συνάντηση, συνεργασία και μη συνεργασία.

ΚΕΦΑΛΑΙΟ 3

ΕΞΑΠΑΤΗΣΗ ΚΑΙ ΕΝΤΟΠΙΣΜΟΣ ΕΞΑΠΑΤΗΣΗΣ ΣΤΑ ΠΟΛΥΠΡΑΚΤΟΡΙΚΑ ΣΥΣΤΗΜΑΤΑ

3.1 Εισαγωγή

Σε αντίθεση με την εμπιστοσύνη που έχουμε ήδη μελετήσει, τα τελευταία χρόνια έχουν γίνει λίγες σχετικά έρευνες για τον τομέα της εξαπάτησης (deception) στα πολυπρακτορικά συστήματα. Όμως η κοινωνική προσέγγιση των υπολογιστών αδιαμφισβήτητα συνεπάγεται και την εξάπλωση της εξαπάτησης (απάτης) στις ψηφιακές κοινωνίες. Συνεπώς είναι απαραίτητη αν όχι αναγκαία η μελέτη αυτού του φαινομένου που ολοένα αυξάνεται στα πολυπρακτορικά συστήματα.

Με την έννοια εξαπάτηση ορίζουμε την κάθε πράξη και τα αποτελέσματα αυτής που αποσκοπούν στην πρόκληση λανθασμένων πεποιθήσεων για ένα συγκεκριμένο γεγονός ή στην απομάκρυνση πρακτόρων από τις γνώσεις που σχετίζονται με την επίτευξη των στόχων τους [19]. Με πιο απλά λόγια, ένας άλλος ορισμός που μπορεί να δοθεί για την εξαπάτηση είναι ο εξής: εξαπάτηση είναι μια ενέργεια ενός πράκτορα που σκοπό έχει την απόκρυψη της αλήθειας, με σκοπό να επωφεληθεί από την ενέργεια αυτή.

Οι υπολογιστές έχουν αναδείξει νέους τρόπους για εξαπάτηση. Η συνήθεια ή η διάθεση που υπάρχει για εξαπάτηση, δυναμώνει με την χρήση υπολογιστών λόγω της ανωνυμίας που υπάρχει στο διαδίκτυο. Πλέον, το αξιοσημείωτο είναι πως φαινόμενα εξαπάτησης δεν παρατηρούνται μόνο μεταξύ των ανθρώπων, αλλά και μεταξύ ανθρώπων με τους πράκτορες, αλλά ακόμη και μεταξύ των πρακτόρων μεταξύ τους. Σε πολλές περιπτώσεις, οι πράκτορες έχουν σχεδιαστεί και προορίζονται γι' αυτόν ακριβώς τον λόγο, για να εξαπατούν.

3.2 Κατηγορίες εξαπάτησης

3.2.1 Κατηγοριοποίηση με βάση την φύση των συμμετεχόντων

Όπως έχουμε ήδη αναφέρει, οι περιπτώσεις εξαπάτησης δεν περιορίζονται πλέον μόνο μεταξύ των ανθρώπων, αλλά και μεταξύ ανθρώπου-πράκτορα και μεταξύ πράκτορα-πράκτορα [19].

Εξαπάτηση πράκτορα-πράκτορα

Η επανάσταση από την κατανεμημένη τεχνητή νοημοσύνη (Distributed Artificial Intelligence), στα πολυπρακτορικά συστήματα (MAS), οδήγησε σε πράκτορες που νοιάζονται μόνο για το δικό τους συμφέρον. Τα DAI συστήματα θεωρούνται κλειστά συστήματα πρακτόρων, πράκτορες των οποίων ο λόγος δημιουργίας είναι να συνεργάζονται και να επιλύουν προβλήματα[20]. Από την άλλη, τα συστήματα MAS έχουν σχεδιαστεί για ανοιχτά συστήματα, συστήματα δηλαδή στα οποία οι πράκτορες εισέρχονται και εξέρχονται σε τυχαίες χρονικές στιγμές. Στα συστήματα αυτά, ο κάθε πράκτορας έχοντας το δικό του συμφέρον, επικοινωνεί, διαπραγματεύεται και τελικά αν είναι προς το συμφέρον του θα συνεργαστεί με τον άλλον πράκτορα. Οι πράκτορες στα συστήματα MAS δεν είναι απαραίτητα καλοπροαίρετοι, συνεργάσιμοι και ειλικρινείς μεταξύ τους.

Εξαπάτηση πράκτορα-χρήστη

Οι πράκτορες που υποστηρίζουν αναζήτηση και ανάκτηση πληροφοριών , ηλεκτρονικό εμπόριο, η οποιουδήποτε τύπου κοινωνικό έλεγχο, είναι πιθανό να εξαπατήσουν τους χρήστες για τους οποίους είναι σχεδιασμένοι να δουλεύουν. Μερικά παραδείγματα εξαπάτησης μεταξύ πράκτορα-χρήστη αναφέρονται παρακάτω.

Συστήματα πληροφοριών και ανάκτησης δεδομένων

Τα συστήματα βάσεων δεδομένων παραπληροφορούν όσους χρήστες δεν έχουν εξουσιοδότηση για να εισέλθουν στα συστήματα αυτά, προκειμένου να προστατέψουν τα απόρρητα δεδομένα τους. Μια γνωστή τεχνική στα συστήματα των βάσεων δεδομένων είναι να δίνονται σκοπίμως δίνονται λανθασμένες πληροφορίες από μεριάς των πρακτόρων, με στόχο τα επίπεδα ασφαλείας να προστατεύονται από τους κακόβουλους χρήστες[21]. Το αξιοσημείωτο στη συγκεκριμένη περίπτωση είναι ότι τα συστήματα δεν αρνούνται να δώσουν πληροφορίες στους χρήστες, άλλα τους παραπληροφορούν, και η παραπληροφόρηση αυτή γίνεται με τρόπο που να εξυπηρετεί τον πράκτορα.

Ηλεκτρονικό εμπόριο

Στην περίπτωση του ηλεκτρονικού εμπορίου, η εξαπάτηση μπορεί να λάβει χώρα σε μια δημοπρασία[22]. Ο πράκτορας λειτουργεί από μεριάς του αγοραστή κακοπροαίρετα για να δημιουργήσει πρόβλημα στη δημοπρασία. Έτσι, ο πράκτορας κάνει μεγαλύτερες προσφορές στον πωλητή ανεβάζοντας την τιμή. Αυτή η τεχνική εξαπάτησης απαιτεί την μη δημοσιοποίηση του κάθε υποψηφίου αγοραστή ,για να μη γίνεται αντιληπτός ο πράκτορας.

Πράκτορες βοηθοί

Οι πράκτορες βοηθοί που δραστηριοποιούνται προς συμφέρον του χρήστη, πολλές φορές μπορεί να χρειαστεί να τον εξαπατήσουν ώστε να τον κάνουν να λάβει μια σωστή απόφαση για ένα ζήτημα, γιατί ο ίδιος ο χρήστης ίσως επέλεγε κάτι διαφορετικό που δε θα ήταν προς όφελος του. Ένα πολύ χαρακτηριστικό παράδειγμα από την καθημερινή μας ζωή, είναι ότι για μία πολύ σοβαρή ασθένεια, ο γιατρός προκειμένου να ενθαρρύνει τον ασθενή και να του ανεβάσει το ηθικό, θα του δώσει πολύ μεγαλύτερο ποσοστό επιτυχίας της θεραπείας από ότι είναι στην πραγματικότητα[23]. Έστω επίσης ότι είχαμε έναν πράκτορα ο οποίος από μόνος του μας οργάνωνε το πρόγραμμα. Στην περίπτωση που ήμασταν πολύ αργοπορημένοι στα ραντεβού μας, ο ίδιος ο πράκτορας έχοντας τη γνώση αυτή, θα μας έλεγε λανθασμένη ώρα συνάντησης (πιο νωρίς), ώστε να φτάσουμε εγκαίρως στο ραντεβού μας. Στις περιπτώσεις αυτές η εξαπάτηση είναι καλοπροαίρετη και έχει σα στόχο να μας βοηθήσει καλύπτοντας δικές μας αδυναμίες.

3.2.2 Κατηγοριοποίηση με βάση τη φύση της εξαπάτησης

Η εξαπάτηση μπορεί να διακριθεί με βάση τα κίνητρα που οδηγούν κάποιον να εξαπατήσει και φυσικά τον τρόπο και την τεχνική που θα ακολουθήσει κάποιος για να προβεί στην ενέργεια αυτή. Η εξαπάτηση μπορεί να χωριστεί σε δύο βασικές κατηγορίες[24]:

Άμεση εξαπάτηση

Με τον όρο άμεση εξαπάτηση εννοούμε ότι ο πράκτορας προκειμένου να εξαπατήσει δημιουργεί ψευδείς πεποιθήσεις και αντιλήψεις στους άλλους πράκτορες, και χωρίζεται σε τρεις υποκατηγορίες:

- **Εξαπάτηση μίμησης (mimicking deception).** Ο πράκτορας που θέλει να εξαπατήσει τον άλλο πράκτορα προσπαθεί να τον κάνει να μιμηθεί λανθασμένες συμπεριφορές.
- **Εξαπάτηση εφεύρεσης (invention deception).** Σ' αυτή την κατηγορία της εξαπάτησης γίνεται μια προσπάθεια δημιουργίας ενός νέου πλαισίου μέσα στο οποίο ο πράκτορας θα δραστηριοποιηθεί. Οι κανόνες και η φύση του νέου αυτού περιβάλλοντος θα είναι τελείως διαφορετικοί από την πραγματικότητα. Έτσι τα ερεθίσματα και οι γνώση που θα έχει ο πράκτορας από ένα λανθασμένο περιβάλλον, θα τον οδηγούν σε λάθος συμπεράσματα και αποφάσεις.
- **Εξαπάτηση παραπλάνησης (decoying deception).** Στην κατηγορία αυτή, γίνεται μια προσπάθεια απόκρυψης του πραγματικού στόχου από τον

πράκτορα. Στη θέση του πραγματικού στόχου, στον πράκτορα δίνεται ένας ψεύτικος στόχος για να επιτύχει που καμία σχέση δεν έχει με τον αρχικό.

Έμμεση Εξαπάτηση

Με τον όρο έμμεση εξαπάτηση εννοούμε την απόκρυψη της αλήθειας με έμμεσο τρόπο και διακρίνεται στις εξής τρεις κατηγορίες:

- **Εξαπάτηση συγκαλύψης (masking deception).** Σε αυτή την περίπτωση η εξαπάτηση γίνεται από ένα πράκτορα, ο οποίος προσπαθεί να αποκρύψει την πραγματικότητα στο περιβάλλον στο οποίο ενεργεί.
- **Εξαπάτηση επανασυσκευασίας (repackaging deception).** Στην περίπτωση αυτή, ο πράκτορας προσπαθεί να κρύψει την αλήθεια με το να την παρουσιάσει με μια τελείως διαφορετική εμφάνιση κάνοντάς την να φαίνεται σα ψέμα.
- **Εξαπάτηση αβεβαιότητας (uncertainty deception).** Στην κατηγορία αυτή συγκαταλέγονται οι περιπτώσεις στις οποίες ο πράκτορας προσπαθεί να δημιουργήσει κλίμα αβεβαιότητας για την πραγματικότητα και με τον τρόπο αυτό οι άλλοι πράκτορες να μην εμπιστεύονται λύσεις που υπάρχουν στο περιβάλλον.

3.3 Εντοπισμός εξαπάτησης

Ένα πολύ σημαντικό ζήτημα που ανακύπτει από τα παραπάνω, είναι κατά πόσο και με ποίο τρόπο μπορούμε να εντοπίζουμε την εξαπάτηση στα πολυπακτορικά συστήματα. Όπως γίνεται εύκολα αντιληπτό, ο σχεδιασμός και η ανάπτυξη ενός τέτοιου μοντέλου είναι απαιτητική δουλειά. Μια τέτοια προσπάθεια παρουσίασαν στην εργασία τους οι Eugene Santos Jr., Deqing Li, Xiuqing Yuan [25]. Τέτοιες έρευνες στο πεδίο των πολυπρακτορικών συστημάτων υπάρχουν ελάχιστες έως και τη στιγμή που γράφεται η εργασία αυτή, και η εξέταση του ζητήματος του εντοπισμού της εξαπάτησης και της εξαπάτησης γενικότερα αποτελεί ένα ενδιαφέρον πεδίο έρευνας. Από κοινωνικής σκοπιάς μόνο το 60% των πρόσωπο-με-πρόσωπο εξαπατήσεων εντοπίζονται [26]. Το ποσοστό αυτό είναι σημαντικά μικρότερο όταν έχουμε να κάνουμε με ηλεκτρονικά μέσα και ηλεκτρονικούς υπολογιστές[27].

Στις επόμενες παραγράφους θα παρουσιαστεί θεωρητικά μια μέθοδος που ακολουθούμε για να εντοπίσουμε τυχόν εξαπατήσεις μεταξύ των πρακτόρων.

3.3.1 Θεωρητικό μοντέλο Εντοπισμού εξαπάτησης Johnson & Grazioli

Οι Johnson και Grazioli [28] προτείνουν μια γενική διαδικασία εντοπισμού της εξαπάτησης που αποτελείται από τέσσερα στάδια.

1^ο Στάδιο. Αρχικά παρατηρούμε τυχόν κακές ενέργειες των πρακτόρων μέσα στο περιβάλλον. Αν εντοπίσουμε κακές επιρροές, εντατικοποιούμε την αναζήτηση για να εστιάσουμε περισσότερο στο σημείο που γίνεται η εξαπάτηση.

2^ο Στάδιο. Έπειτα, προσπαθούμε να εντοπίσουμε πιο συγκεκριμένα τα σημεία και τους πράκτορες που παρατηρείται υποψία χειραγώγησης.

3^ο Στάδιο. Για τα σημεία για τα οποία θεωρούμε ότι έχουμε εξαπάτηση, γίνεται εκτίμηση του είδους και του μεγέθους της εξαπάτησης, και συγκρίνουμε αν αυτή η εξαπάτηση έχει χαρακτηριστικά ίδια με αυτά που παρατηρήσαμε στην αρχή.

4^ο Στάδιο. Συνδυάζοντας όλες τις ύποπτες για εξαπάτηση υποθέσεις, εξάγεται ένα τελικό συμπέρασμα για τη φύση και την έκταση της εξαπάτησης.

3.3.2 Ολοκληρωμένο μοντέλο εντοπισμού εξαπάτησης

Η βασική ιδέα για τον εντοπισμό της απάτης στα πολυπρακτορικά συστήματα είναι ότι αν ένας αριθμός πρακτόρων έχει την ίδια γνώση, τα συμπεράσματα στα οποία θα καταλήξουν οι πράκτορες πρέπει να είναι παρόμοια. Σε περίπτωση που το συμπέρασμα ενός πράκτορα αποκλίνει πολύ από τα αναμενόμενα, τότε είναι πιθανό ο πράκτορας να έχει εξαπατήσει.

Υπάρχουν τέσσερις μέθοδοι που λαμβάνουν χώρα στη μέθοδο εντοπισμού της απάτης:

- Αρχικά, υποθέτουμε ότι όλοι οι πράκτορες του συστήματος μοιράζονται το ίδιο υπόβαθρο γνώσης και έχουν παρόμοια δομή, οπότε για να προβλέψουμε την άποψη ενός πράκτορα για ένα συγκεκριμένο ζήτημα αρκεί να γνωρίζουμε την άποψη των υπολοίπων πρακτόρων. Όμως για να προβλέψουμε την γνώμη ενός πράκτορα, πρέπει να λάβουμε υπόψη και τις σχέσεις που έχει ο πράκτορας αυτός με τους άλλους πράκτορες του συστήματος. Στην ουσία, η πρώτη διαδικασία είναι η σύγκριση πρακτόρων. Για να παρατηρήσουμε τη συμπεριφορά των πρακτόρων και τις σχέσεις που αναπτύσσουν μεταξύ τους, οι πράκτορες αρχικά υπόκεινται σε κάποια τεστ.

- Έπειτα, χρησιμοποιούμε όλες τις απόψεις των πρακτόρων από την παραπάνω διαδικασία για να υπολογίσουμε τις συσχετίσεις Pearson [29] που έχουν μεταξύ τους. Οι τιμές των συσχετίσεων δείχνουν τη γραμμική εξάρτηση μεταξύ των πρακτόρων.
- Ακολούθως, συνεχίζουμε με την διαδικασία της πρόβλεψης, χρησιμοποιώντας την τεχνική πρόβλεψης GroupLens [30]. Με την τεχνική αυτή μπορούμε να προβλέψουμε την αντίδραση ενός πράκτορα βασισμένοι στην συσχέτιση του πράκτορα αυτού με όλους τους άλλους πράκτορες.
- Η τελική διαδικασία είναι να εντοπίσουμε τυχόν απάτες στο σύστημα, βάσει του πόσο μακριά είναι η άποψη ενός πράκτορα από την άποψη των υπολοίπων πρακτόρων. Στην πράξη, υποθέτουμε ότι το σφάλμα της άποψης του πράκτορα και της πραγματικής άποψης ακολουθεί κανονική κατανομή. Αν η διαφορά της πραγματικής άποψης από την άποψη του πράκτορα είναι μεγαλύτερα από το τριπλάσιο της τυπικής απόκλισης (standard deviation), τότε θεωρούμε ότι ο πράκτορας μάλλον έχει εξαπατήσει. Παρατηρώντας την κανονική κατανομή το ποσοστό αυτό ανέρχεται στο 0.3% των συνολικών παρατηρήσεων.

Η αποδοτικότητα αυτής της μεθόδου φαίνεται σε πολλές μελέτες. Μια τέτοια μελέτη έγινε προκειμένου το μοντέλο να παρακολουθήσει μια θεραπεία ασθενών [25]. Οι παράμετροι που εισήχθησαν στο πείραμα είναι ο αριθμός των πρακτόρων, την perturbation value, που είναι η τιμή που ορίζει την ομοιότητα των απόψεων των πρακτόρων, ο αριθμός των μεταβλητών που χρησιμοποιούνται ως ενδείξεις για την ασθένεια και ο αριθμός των τυπικών αποκλίσεων (STD) για τις οποίες η διαφοράς μεταξύ των προβλεπόμενων γνώμων από τις πραγματικές να είναι αποδεκτές. Για να εξετάσουμε την επίδραση αυτών των παραμέτρων όπως είπαμε παραπάνω, γίνονται κάποια τεστ όπου μεταβάλλουμε μια εκ των παραμέτρων αυτών κρατώντας σταθερές τις υπόλοιπες. Κάθε φορά αυτό που λαμβάνουμε σαν αποτέλεσμα αυτών των μεταβολών είναι η τιμή της μέσης ανίχνευσης της εξαπάτησης (mean deception detection). Οι τιμές των mean deception detection παρουσιάζονται στους παρακάτω πίνακες.

Στο πρώτο πείραμα μεταβάλλεται ο αριθμός των πρακτόρων και ταυτόχρονα μεταβάλλονται και τα όρια του perturbation value. Είναι προφανές ότι καθώς τα όρια του perturbation value αυξάνονται, οι τιμές του mean deception detection μειώνονται καθώς οι απόψεις των πρακτόρων γίνονται λιγότερο ομοιογενείς (περισσότερος θόρυβος στο σύστημα), και συνεπώς τυχόν εξαπάτηση γίνεται δυσκολότερα αντιληπτή. Από τις τιμές του πίνακα, φαίνεται ότι ο αριθμός των

πρακτόρων στο σύστημα δεν επηρεάζει σημαντικά τις τιμές του mean deception detection.

Pert. Interv. Πράκτ.	(-0.1,0.1)	(-0.2,0.2)	(-0.3,0.3)	(-0.4,0.4)
3	0.936	0.7843	0.7285	0.6088
10	0.9022	0.7433	0.6261	0.5878
30	0.8902	0.7455	0.6396	0.5692
100	0.8923	0.747	0.637	0.5792

Στο δεύτερο πείραμα μεταβάλλεται ο αριθμός των ενδείξεων. Όπως είναι λογικό, όσο περισσότερες ενδείξεις έχουμε, τόσο πιο εύκολα εντοπίζουμε την εξαπάτηση στο σύστημα.

Pert. Interv. Ενδείξεις	(-0.1,0.1)	(-0.2,0.2)	(-0.3,0.3)	(-0.4,0.4)
1-10	0.9066	0.7715	0.638	0.5568
11-20	0.955	0.8922	0.8404	0.7723
21-30	0.9829	0.9716	0.9435	0.9359

Στο τρίτο πείραμα μεταβάλλεται ο αριθμός των τυπικών αποκλίσεων. Όσο περισσότερες τυπικές αποκλίσεις έχουμε, τόσο μεγαλύτερο επιτρέπουμε να είναι το σφάλμα, και συνεπώς η εξαπάτηση γίνεται λιγότερο αντιληπτή.

Pert. Interv. #STD	(-0.1,0.1)	(-0.2,0.2)	(-0.3,0.3)	(-0.4,0.4)
4	0.8496	0.6403	0.5195	0.4349
3	0.9037	0.7748	0.6163	0.5327
2	0.9383	0.8529	0.7445	0.6858
1	0.9885	0.9556	0.9181	0.859

Τέλος, τα αποτελέσματα των παραπάνω πινάκων, μέσω της ανάλυσης ANOVA [31] παράγουν τον παρακάτω πίνακα. Η ανάλυση ANOVA (analysis of variance), είναι μια στατιστική μέθοδος που χρησιμοποιείται για να υπολογίσει διαφορές μεταξύ δειγμάτων για διαφορετικές μεταβλητές του συστήματος. Στην περίπτωση μας, να συγκρίνουμε τέσσερα γκρουπ, στο ένα μεταβάλλεται το πλήθος των πρακτόρων, στο άλλο αλλάζουν τα perturbations value, στο άλλο ο αριθμός των ενδείξεων, και στο τελευταίο ο αριθμός των τυπικών αποκλίσεων. Από την ανάλυση ANOVA προκύπτουν οι τιμές F value, F critical value και P value για κάθε παράμετρο. Αν η τιμή της F value είναι πιο μεγάλη από την τιμή της F critical value, τότε αυτό σημαίνει ότι η παράμετρος επηρεάζει σημαντικά την ανίχνευση της εξαπάτησης, σε διαφορετική περίπτωση, η παράμετρος δεν παίζει σπουδαίο ρόλο στην ανίχνευση της εξαπάτησης. Από τον πίνακα που ακολουθεί διαπιστώνουμε ότι το perturbation value, ο αριθμός των ενδείξεων και ο αριθμός των τυπικών αποκλίσεων επηρεάζουν σημαντικά την ανίχνευση της εξαπάτησης, ενώ το πλήθος των πρακτόρων όχι και τόσο.

Parameter	Agents	Perturbation	Evidence	#STD
F value	2.5	346.3	1524.04	1845.03
F critical	3.1	2.68	3.1	2.68
Significance	NOT	YES	YES	YES

3.4 Συμπεράσματα

Οι αναφορές στον τομέα της εξαπάτησης των πολυπρακτορικών συστημάτων είναι σχετικά πρόσφατες (μετά το 2000). Αυτό που με σιγουριά μπορούμε να πούμε είναι ότι θα αποτελέσει ένα από τα πιο πολυσυζητημένα θέματα σε επίπεδο έρευνας και κατ' επέκταση εφαρμογής. Αυτό, συμβαίνει γιατί οι πράκτορες έχουν εισέλθει σχεδόν σε όλες τις πτυχές της καθημερινής ζωής. Πράκτορες χρησιμοποιούνται για την πρόβλεψη του καιρού, για την πρόβλεψη μιας ασθένειας, στο ηλεκτρονικό εμπόριο κλπ. Οπότε η ανάγκη για νέα ακριβή και αποδοτικά μοντέλα εντοπισμού της εξαπάτησης ολοένα και θα αυξάνεται. Ένα σχετικά αξιόπιστο μοντέλο εντοπισμού εξαπάτησης είναι αυτό που επιγραμματικά αναφέρεται στο κεφάλαιο 3.3.2. Η βασική ιδέα για τον εντοπισμό της εξαπάτησης είναι ότι η συμπεριφορά των πρακτόρων πρέπει να είναι παραπλήσια, αν οι πράκτορες μοιράζονται την ίδια γνώση. Οπότε, όποιος πράκτορας αποκλίνει εξετάζεται περαιτέρω για τυχόν εξαπάτηση. Πρακτικά, το μοντέλο χρησιμοποιεί τεχνικές αρχικά για να υπολογίσει

τις συσχετίσεις που έχουνε οι πράκτορες μεταξύ τους ,και έπειτα για να προβλέψουμε τις αντιδράσεις των πρακτόρων σε θεωρητικά σενάρια. Στα θετικά του μοντέλου είναι η δυνατότητα εφαρμογής του σε πληθώρα εφαρμογών και η εφαρμογή ακριβών μοντέλων συσχέτισης και πρόβλεψης. Αυτή η μέθοδος παρουσιάζει και ελλείψεις. Αρχικά, όλες οι παράμετροι αντιμετωπίζονται ως εξίσου σημαντικές, κάτι που δεν ισχύει πάντα στα πραγματικά πολυπρακτορικά συστήματα. Επίσης, η δομή του συστήματος είναι πολύ ιδανική σε σύγκριση με την πραγματικότητα, όπως για παράδειγμα ότι όλοι οι πράκτορες έχουν την ίδια βάση γνώσης. Ωστόσο, παρ' όλα τα αρνητικά της σημεία, η μέθοδος αυτή είναι αρκετά αποδοτική στον εντοπισμό της εξαπάτησης στα πολυπρακτορικά συστήματα.

ΚΕΦΑΛΑΙΟ 4

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η περιοχή της εμπιστοσύνης και ιδιαίτερα της εξαπάτησης συνεχώς θα αναπτύσσεται. Η ανάγκη για ασφαλείς συναλλαγές μεταξύ των πρακτόρων είναι μια πραγματικότητα που θα δημιουργεί την ανάγκη για νέα μοντέλα εμπιστοσύνης.

Αυτό που παρατηρούμε με τα μοντέλα εμπιστοσύνης είναι ότι δεν υπάρχει κάποιος που να υπερτερεί γενικά έναντι των άλλων. Κάθε ένα από αυτά είναι πιο αποδοτικό στο περιβάλλον για το οποίο προορίζεται. Αρχικά για κάθε μοντέλο εμπιστοσύνης γίνονται κάποιες παραδοχές που λαμβάνουν υπόψη τα ιδιαίτερα χαρακτηριστικά του προβλήματος και κάνουν τους υπολογισμούς της εμπιστοσύνης ευκολότερους. Έπειτα, ακολουθεί η διαδικασία υπολογισμού της εμπιστοσύνης όπου εφαρμόζονται οι κατάλληλοι μαθηματικοί υπολογισμοί και η απαιτούμενη θεωρία πιθανοτήτων.

Στις επιστημονικές εργασίες για τον υπολογισμό της εμπιστοσύνης, η ασάφεια των μοντέλων εμπιστοσύνης είναι ένα αρκετά συνηθισμένο φαινόμενο. Οι συγγραφείς, τις περισσότερες φορές, εστιάζουν την προσοχή τους και προτείνουν λύσεις για μια μικρή περιοχή του προβλήματος. Σύμφωνα με αυτό και σε συνδυασμό με τις παραδοχές που γίνονται, δημιουργείται η αμφιβολία για το κατά πόσο ένα μοντέλο είναι εφαρμόσιμο και πρακτικό.

Όπως έχουμε ήδη πει, τα πεδία της εμπιστοσύνης και της εξαπάτησης έχουν μεγάλο περιθώριο για νέα έρευνα και υπάρχει μεγάλο εύρος έρευνας για νέους ερευνητές που θέλουν με αυτούς τους τομείς. Όσον αφορά την εμπιστοσύνη, θα ήταν πολύ χρήσιμο να δημιουργηθούν ειδικές πλατφόρμες αξιολόγησης των μοντέλων εμπιστοσύνης. Με την σχεδίαση τέτοιων εργαλείων, θα είναι δυνατή η αξιολόγηση της εμπιστοσύνης για κάθε μοντέλο που ο χρήστης επιθυμεί. Στον τομέα της εξαπάτησης, μια μελλοντική δουρεία μπορεί να είναι η δημιουργία ενός νέου μοντέλου υπολογισμού της εξαπάτησης. Μέχρι την συγγραφή της παρούσας διπλωματικής εργασίας εντοπίσαμε μόνο ένα ολοκληρωμένο μοντέλο εντοπισμού της εξαπάτησης, οπότε σίγουρα η ενασχόληση με τη δημιουργία ενός τέτοιου μοντέλου θα έχει ιδιαίτερο ενδιαφέρον και ίσως καλύψει αδυναμίες του αρχικού μοντέλου αναβαθμίζοντάς το.

Από την ενασχόλησή μου με τα αντικείμενα της εμπιστοσύνης και της εξαπάτησης στο πλαίσιο της παρούσας διπλωματικής εργασίας, αποκόμισα εμπειρία και γνώση που σίγουρα θα αποτελέσουν χρήσιμα εφόδια για το μέλλον μου. Το ότι τα αντικείμενα αυτά είναι πολύ πρόσφατα και σε συνδυασμό με το ότι η τεχνητή

νοημοσύνη κατακλύζει την ζωή μας, η ερευνητική ενασχόληση οποιουδήποτε νέου επιστήμονα με τους τομείς αυτούς είναι μια πρόκληση και μια πρωτοποριακά δημιουργική δραστηριότητα. Κλείνοντας, θα με χαροποιούσε ιδιαίτερα αν η εργασία αυτή μελλοντικά βοηθούσε κάποιον συνάδελφο φοιτητή και τον διευκόλυνε στη δική του έρευνα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] S. Marsh, "Formalizing Trust as a Computational Concept", PhD thesis, University of Stirling, U.K., 1994.
- [2] A.Josang, C.Keser and T.Dimitrakos, "Can We Manage Trust ?", In the Proceedings of the 3rd International Conference on Trust Management, (iTrust), Paris, May 2005.
- [3] D. Gambetta. "Can We Trust Trust?", In D. Gambetta, editor, Trust: Making and Breaking Cooperative Relations, pages 213-238. Basil Blackwell. Oxford, 1990.
- [4] T.Grandison and M.Sloman, "A Survey of Trust in Internet Applications", IEEE Communications Surveys, 2000.
- [5] Q.Z hang, T.Yu and K.Irwin, "A Classification Scheme for Trust Functions in Reputation-Based Trust Management", In ISWC'04, 3rd International Semantic Web Conference, Workshop on Trust, Security, and Reputation on the Semantic Web", Hiroshima, Japan, 2004.
- [6] Μιχάλης Μπαλταγιάννης, "Αναπαράσταση και Διαχείριση Εμπιστοσύνης σε Ανοιχτά Πολυπρακτορικά Συστήματα", Μεταπτυχιακή Διατριβή, Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών και Δικτύων, Πανεπιστήμιο Θεσσαλίας, 2006.
- [7] eBay, <http://www.eBay.com>
- [8] <https://www.feedbackfive.com/blog/amazon-feedback-vs-ebay-feedback/>
- [9] Amazon, <http://www.amazon.com>
- [10] T. Dong Huynh, Nicholas R. Jennings and Nigel R. Shadbolt, "An integrated trust and reputation model for open multi-agent systems", Auton Agent Multi-Agent Sys 13: 119–154, Springer Science+Business Media, 2006.
- [11] T. D. Huynh, N. R. Jennings and N. R. Shadbolt, "Developing an integrated trust and reputation model for open multi-agent systems", in Proc. 7th Int Workshop on Trust in Agent Societies, 2004..

- [12] Babak Khosravifar, "Trust and Reputation in Multi-Agent Systems", A Thesis in The Department of Electrical and Computer Engineering, Concordia University, 2012
- [13] Manh Hung Nguyen and Dinh Que Tran, "A combination trust model for multi-Agent systems", *International Journal of Innovative Computing, Information and Control*, Volume 9, Number 6, June 2013
- [14] L.Xi ong and L.Liu, "Peer Trust: A trust mechanism for an open peer-to-peer information system", Technical Report GIT-CC-02-29, Georgia Institute of Technology, 2002
- [15] Lik Mui, Mojdeh Mohtashemi and Ari Halberstadt, "A Computational Model of Trust and Reputation", *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2002
- [16] Στρατής Κούνιας, Χρόνης Μωυσιάδης, "Θεωρία Πιθανοτήτων Ι", Εκδόσεις ΖΗΤΗ, Θεσσαλονίκη 1999.
- [17] Χαράλαμπος Χανιαλίδης, "Συμπερασματολογία και επιλογή μεταβλητών για μοντέλα ποσοστημοριακής παλινδρόμησης", Διπλωματική Εργασία, Πανεπιστήμιο Αθηνών, Τμήμα Μαθηματικών, Μεταπτυχιακό Στατιστικής και Επιχειρησιακής Έρευνας, Οκτώβριος 2010.
- [18] Δημήτριος Ν. Σαββάτης, "Φράγματα τύπου Chernoff και Εφαρμογές", Μεταπτυχιακή διπλωματική εργασία, Πανεπιστήμιο Πατρών, Τμήμα Μαθηματικών, Δεκέμβριος 2015.
- [19] *Cristiano Castelfranchi and Yao-Hua Tan*, "The Role of Trust and Deception in Virtual Societies", *International Journal of Electronic Commerce*, Vol. 6, No. 3, pp. 55–70, Spring 2002
- [20] O'Hare, G, Jennings, "Foundations of Distributed AI", New York: John Wiley, 1996.
- [21] Wagner G.,P. Kandzia and M. Klusch, "Cooperative Information Agents" Berlin: Springer Verlag, , pp. 272–285,1997.
- [22] Rosis F., Grasso F. and Berry D., "Refining medical explanation generation after evaluation", *Artificial Intelligence in Medicine*, 17, 1999.
- [23] Witte K., "The manipulative nature of health communication research. Ethical issues and guidelines", *American Behavioral Scientist*, 38, 2, 1994.
- [24] Whaley B. and Bell J., "Cheating and Deception", Transaction Publishers,1991.

- [25] Eugene Santos Jr., Deqing Li and Xiuqing Yuan, "On Deceptopn Detection in Multi-Agent Systems and Deception Intent", Thayer School of Engineering , Dartmouth, Hanover, New Hampshire, USA, 2008.
- [26] Depaulo B., Kirkendol S., Tang, J. and O'Brien, T., "The motivational impairment effect in the communication of deception: Replications and extensions, " Journal of nonverbalBehavior, 12, 177-202 ,1988.
- [27] George J., Biros D., Burgoon J. and Nunamaker J. Jr., "Training Professionals to Detect Deception", Springer Berlin/Heidelberg, Intelligence and Security Informatics: First NSF/NIJ Symposium, Tucson, AZ, USA, June 2-3 2003.
- [28] Johnson P., Grazioli S., Jamal K. and Berryman R., "Detecting deception: adversarial problem solving in a low base-rate world", Cognitive Science 25 2001.
- [29] <https://thalis.math.upatras.gr/~adk/lectures/ida/lab6/slides6.pdf>
- [30] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl, "Explaining Collaborative Filtering Recommendations", Dept. of Computer Science and Engineering, University of Minnesota, 2010.
- [31] <https://www.spss-tutorials.com/anova-what-is-it/>