



FreeBSD: Ανάπτυξη λογισμικού σε παγκόσμια κλίμακα

Ενα από τα χαρακτηριστικά του ανοικτού κώδικα είναι η παγκοσμιοποίηση της ανάπτυξής του. Σε αυτό το άρθρο τεκμηριώνονται με ποσοτικά δεδομένα οι πρακτικές, οι προκλήσεις που αντιμετωπίζουν οι προγραμματιστές στην ανάπτυξη ενός μεγάλου έργου, όπως το FreeBSD, καθώς και τα αποτελέσματα που προκύπτουν.

ΤΟΥ ΔΙΟΜΗΔΗ ΣΠΙΝΕΛΛΗ *

Αν. Καθηγητή
Τμήμα Διοικητικής
Επιστήμης και Τεχνολογίας
Οικονομικό Πανεπιστήμιο
Αθηνών

<http://www.dmst.aueb.gr/ddss>



Tο FreeBSD (www.freebsd.org) είναι ένα εξελιγμένο λειτουργικό σύστημα που διατίθεται για διάφορες αρχιτεκτονικές υπολογιστών. Σε αντίθεση με το Linux που είναι μόνο πιρήνας και χρειάζεται άλλα προγράμματα για να δουλέψει, το FreeBSD είναι ένα πλήρες λειτουργικό σύστημα. Πρέρχεται από το BSD Unix, μία έκδοση του Unix που αναπτύχθηκε στο Πανεπιστήμιο της Καλιφόρνια, στο Berkeley των ΗΠΑ. Το FreeBSD είναι γνωστό για τη σταθε-

ρότητα και την αξιοπιστία του, γι' αυτό χρησιμοποιείται σε μεγάλα sites, όπως το Yahoo! και εταιφρείς φιλοξενίας ιστοσελίδων, όπως η Host Department. Μάλιστα, η βάση του Mac OS X (του λειτουργικού των υπολογιστών της Apple) προέρχεται από το FreeBSD.

Το μοντέλο ανάπτυξης του FreeBSD είναι παγκόσμιο, δηλαδή, συνεισφέρουν προγραμματιστές από κάθε γωνιά της Γης. Σε αυτό το άρθρο θα δούμε τα αποτελέσματα του μοντέλου στην ποιότητα και στην παραγωγικότητα, διερευνώντας τον τρόπο με τον οποίο η γεωγραφική απόσταση επηρεάζει τη συνερ-

γασία των μελών του έργου. Ειδικότερα, θα ασχοληθούμε με τις πρακτικές που υιοθετεί ένας μεγάλος αριθμός από προγραμματιστές και γι' αυτό θα χρησιμοποιήσουμε δεδομένα από το σύστημα διαχείρισης του έργου, τον πηγαίο κώδικα που παράγεται, τις γεωγραφικές συντεταγμένες των προγραμματιστών και το σύστημα αναφοράς προβλημάτων.

Το μοντέλο ανάπτυξης

Το FreeBSD αναπτύσσεται και διανέμεται ως λογισμικό ανοικτού κώδικα από μία ομάδα με πάνω από 350 μέλη που βρίσκονται σε διάφορα μέρη του κό-

*Απόδοση από το Δημήτρη Καλαμαρά της πρωτότυπης δημοσίευσης "Global Software Development in the FreeBSD Project" στο συνέδριο "International Workshop on Global Software Development for the Practitioner", ACM Press, σελ. 73-79, Μάιος 2006.

ομου. Η δουλειά τους μπορεί να ταξινομήσει γενικά σε συγγραφή κώδικα για τον πυρήνα του συστήματος και τα διάφορα εργαλεία του, τη μεταφορά σε αυτό εφαρμογών από άλλα λειτουργικά και, φυσικά, την τεκμηρίωση. Ο συντονισμός αυτής της παγκόσμιας προσπάθειας γίνεται μέσω διάφορων υποδομών, που περιγράφω συνοπτικά παρακάτω.

Κατ' αρχάς, υπάρχει ένα αποθετήριο (repository-ένα είδος διαδικτυακής αποθήκης) που στεγάζει το "ούστημα διαχείρισης σχηματισμών" και βασίζεται στο CVS (www.nongnu.org/cvs) . Σε αυτό το σύστημα υπάρχει πάντοτε η τελευταία έκδοση όλου του κώδικα και της τεκμηρίωσης του project, οι κλάδοι (branches) συντήρησης παλαιότερων εκδόσεων καθώς και ιστορικό αρχείο μίας δεκαετίας. Ολό το αποθετήριο CVS είναι δημόσια προσβάσιμο από μία διαδικτυακή διεπαφή και ο καθένας μπορεί να κατεβάσει ό, τι αρχεία θέλει.

Υπάρχει ακόμη μία βάση δεδομένων με αναφορές προβλημάτων (που στρέζεται στο σύστημα GNATS), όπου καταγράφονται περιγραφές όλων των ανοικτών ή λυμένων θεμάτων, οι προγραμματιστές που τα χειρίζονται και λεπτομερεις για το τι έχει γίνει. Επίσης, πρέπει να αναφέρουμε το σύστημα tinderbox, το οποίο μεταγλωτίζει συνέχεια το σύστημα από τον τρέχοντα πηγαίο κώδικα του FreeBSD, με σκοπό τον έγκαιρο εντοπισμό προβλημάτων στο αποθετήριο.

Στις υποδομές του έργου συμπεριλαμβάνονται και πάνω από 100 λίστες ηλεκτρονικού ταχυδρομείου (δηλιμόσιας ή ιδιωτικής πρόσβασης), που καλύπτουν διάφορα θέματα και λειτουργούν ως μηχανισμός επικοινωνίας των μελών και των χρηστών του. Στον ιστότοπο του FreeBSD υπάρχει το "Developer's Handbook" (τεκμηρίωση για προγραμματιστές), ενημερωμένες τεχνικές πληροφορίες για κάθε έκδοση, μία έκδοση του αποθετήριου CVS όπου μπορεί να περιηγηθεί κάποιος με το φύλλομετρητή, τα αρχεία των λιστών αλληλογραφίας και μία διεπαφή για την ανάγνωση (μόνο) της βάσης δεδομένων με τις αναφορές των προβλημάτων. Τέλος, ένα δίκτυο υπολογιστών, προσβάσιμο σε κάθε μέλος του έργου μέσω του Διαδικτύου, παρέχει στους προγραμματιστές έναν κοινό χώρο εργασίας για τη μεταγλώπιση και δοκιμή του κώδικα τους σε διάφορες αρχιτεκτονικές.

Τα μέλη

Κάθε μέλος της ομάδας ανάπτυξης του FreeBSD που συνεισφέρει με κώδικα ή τεκμηρίωση ονομάζεται διαπράκτης (committer) και οι συνεισφορές του (μέσω του συστήματος ελέγχου σχηματισμών CVS) λέγονται διαπράξεις (commits). Η ανάδειξη του γίνεται κατόπιν ψηφοφο-

ρίας μετά από πρόταση ενός άλλου διαπράκτη. Στη συνέχεια, ακολουθεί μία εκπαιδευτική περίοδος (mentorship), όπου ένα μέλος, συνήθως ο ίδιος που τον πρότεινε, παίζει το ρόλο του μέντορα του νέου διαπράκτη και εγκρίνει τις συνεισφορές του. Τα μέλη της ομάδας ανάπτυξης συνήθως ξεκινούν συνεισφέροντας κώδικα ανεξάρτητα, μέχρι που κάποιο μέλος του FreeBSD θα ενδιαφέρεται για τη δουλειά τους και θα τους προτείνει να γίνουν επίσημα διαπράκτες.

Τα μέλη της ομάδας ανάπτυξης του FreeBSD είναι κυρίως εθελοντές, αν και υπάρχουν εταιρείες που επιπρέπουν ή βάζουν κάποιους εργαζόμενούς τους να συνεισφέρουν ως μέρος της δουλειάς τους. Το έργο FreeBSD διοικείται από μία εκλεγμένη ομάδα ("Core Team") που αποφασίζει για τους στόχους και την κατεύθυνσή του, εγκρίνει τα νέα μέλη και επιλέγει τις διαφορές που ανακύπτουν.

Για τα διάφορα μέρη του έργου (ασφά-

γώνει πριν από μία νέα έκδοση χωρίς να έχουν πάρει άδεια από την ομάδα που είναι υπεύθυνη για την κυκλοφορία των νέων εκδόσεων. Επίσης, αποδοκιμάζονται οι μεγάλες αλλαγές σε κώδικα που συνητρεπτείται από άλλον προγραμματιστή ή οι αλλαγές που αναιρούν τη δουλειά ενός άλλου μέλους.

Μεθοδολογία

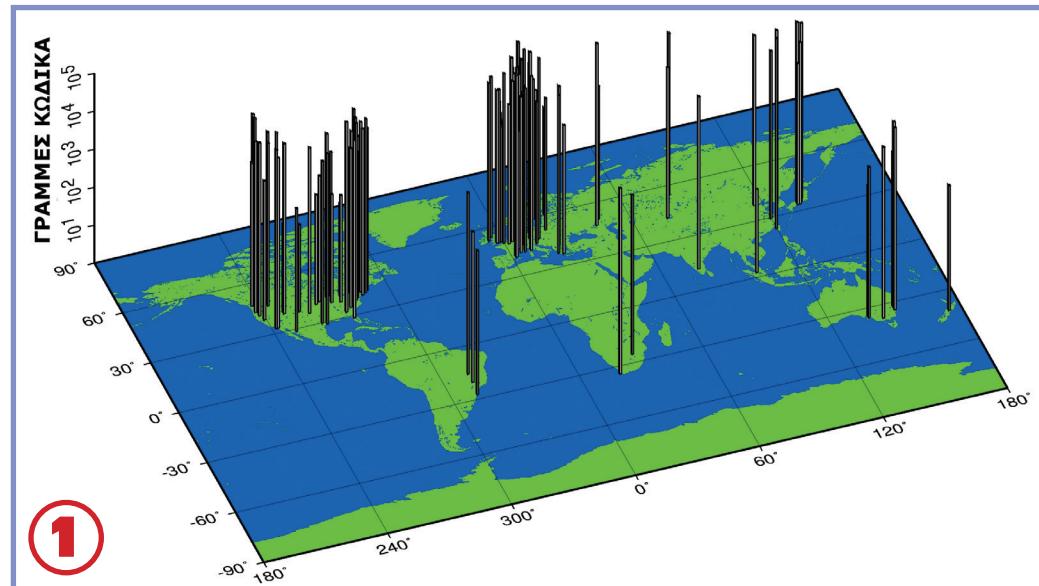
Ας δούμε τώρα πώς θα αναλύσουμε το παγκόσμιο μοντέλο ανάπτυξης του FreeBSD. Οπως είπαμε, τα δεδομένα που προέρχονται από το αποθετήριο CVS του έργου, τη βάση δεδομένων με τις αναφορές προβλημάτων και τις γεωγραφικές συντεταγμένες των προγραμματιστών.

Τα δεδομένα του αποθετηρίου προέρχονται από ένα στιγμιότυπο που πήραμε στις 9 Νοεμβρίου 2005 και εξετάσαμε με τον πελάτη CVS και μερικά εργαλεία του Unix, όπως τα sed, awk, sort

ΟΙ ΠΡΟΓΡΑΜΜΑΤΙΣΤΕΣ ΤΟΥ FREEBSD ΕΙΝΑΙ ΕΘΕΛΟΝΤΕΣ, ΆΛΛΑ ΥΠΑΡΧΟΥΝ ΕΤΑΙΡΕΙΕΣ ΠΟΥ ΕΠΙΤΡΕΠΟΥΝ ΣΕ ΕΡΓΑΖΟΜΕΝΟΥΣ ΤΟΥΣ ΝΑ ΣΥΝΕΙΣΦΕΡΟΥΝ.

λεια, δωρεές, μεταφορές εφαρμογών και κυκλοφορία νέων εκδόσεων) υπάρχουν έχειωριστές ομάδες εργασίας. Οι κανονικοί διαπράκτες έχουν το δικαίωμα τροποποίησης κάθε μέρους του FreeBSD ("προνόμιο διάπραξης") αρκεί να τηρούν μία επίσημη και ανεπίσημη επικέτα. Για παράδειγμα, οι προγραμματιστές δεν μπορούν να υποβάλλουν αλλαγές στην περίοδο που ο κώδικας πα-

και join. Το αποθετήριο περιείχε 19 εκατομμύρια γραμμές κώδικα και τεκμηρώσης, μαζί με 938.000 διαπράξεις από 475 διαφορετικά μέλη που υπεβλήθησαν σε μία περίοδο μίας δεκαετίας περίπου. Τέλος, η βάση δεδομένων με τις αναφορές προβλημάτων (που τηρούνται σε μορφή GNATS) εξερευνήθηκε το Δεκέμβριο του 2005 και εκείνη τη στιγμή περιείχε 90.000 αναφορές.



Γραμμές κώδικα ανά περιοχή της Γης. Η Βόρεια Αμερική και η Ευρώπη έχουν τα πρωτεία στην ανάπτυξη του FreeBSD, ενώ υπάρχουν ομάδες στην Ασία, στην Αυστραλία, στη Νότια Αφρική και στη Νότια Αμερική.

Οι διαπράξεις στο CVS και οι αναφορές GNATS συνδεύονται από επικέτες που δηλώνουν το όνομα του λογαριασμού του αντίστοιχου προγραμματιστή. Αντιστοιχίζοντας τους προγραμματιστές στις τοποθεσίες τους πάνω στη Γη, μπορούμε να εξάγουμε χρονικά συμπεράσματα. Για να το πετύχουμε αυτό, χρησιμοποιήσαμε ένα αρχείο που διανέμεται μαζί με την έκδοση της εφαρμογής XEarth για FreeBSD, το οποίο περιέχει τις γεωγραφικές συντεταγμένες, το όνομα λογαριασμού και την τοποθεσία πολλών από τους προγραμματιστές του FreeBSD. Επιπλέον, μετά από επικοινωνία με την κοινότητα ανάπτυξης, καταφέραμε να αυξήσουμε τον αριθμό των διαπράξεων που μπορούν να αποδοθούν σε κάποια συγκεκριμένη γεωγραφική περιοχή από 71% που ήταν στις 9 Νοεμβρίου στο 79% στις 9 Ιανουαρίου 2006.

να υπολογίσουμε τις αποστάσεις μεταξύ των προγραμματιστών. Για το σκοπό αυτό, χρησιμοποιήθηκε ο τύπος του σφαιρικού νόμου των συνημίτονων που δίνει την απόσταση d μεταξύ δύο σημείων με γεωγραφικό πλάτος δ και μήκος λ πάνω σε μία σφαίρα με ακτίνα R :

$$d = R \cdot \sin^{-1} [\eta \mu \delta_1 \cdot \eta \mu \delta_2 + \sin \delta_1 \sin \delta_2 \cos(\lambda_2 - \lambda_1)]$$

Σημειώνουμε ότι ο παραπάνω τύπος δεν λαμβάνει υπόψη την ελλειψειδή μορφή της Γης, αλλά είναι αρκετά ακριβής για όσα θα κάνουμε παρακάτω.

Παγκόσμια ανάπτυξη

Ενα από τα σημαντικότερα ερωτήματα σχετικά με τις μεθοδολογίες παγκόσμιας ανάπτυξης λογισμικού είναι κατά πόσον αυτές βοηθούν στην πράξη. Το FreeBSD είναι ένα επιτυχημένο έργο που εφαρμόζει μία τέτοια μεθοδολογία και μελετώντας το μπορούμε να βγά-

θεσία (στην ουσία οι προγραμματιστές που ζουν εκεί) απεικονίζεται με τις κάθετες μπάρες στην εικόνα 1. Οπως φαίνεται, λοιπόν, οι περισσότεροι διαπράξτες ζουν στη Βόρεια Αμερική και στην Ευρώπη, ενώ μερικές ομάδες εμφανίζονται στην Ασία, στην Αυστραλία, στη Νότιο Αφρική και στη Νότιο Αμερική.

Ειδίκευση ανά περιοχή

Μία ενδιαφέρουσα ερώτηση που πιθανώς να θέσει κάποιος αφορά στο αν κάποιες περιοχές του κόσμου ειδικεύονται σε διαίτερες δραστηριότητες του έργου, δηλαδή, αν μία γεωγραφική περιοχή επικεντρώνεται σε ένα συγκεκριμένο τμήμα του συστήματος ή προσφέρει ένα συγκεκριμένο είδος υπηρεσίας. Δεδομένης της φύσης του FreeBSD ως εθελοντικού εγχειρήματος, όπου οι συμμετέχοντες μπορούν να συνεισφέρουν όπου και ότι θέλουν, η πιθανή διαίρεση της προσπάθειας σε διάφορες περιοχές θα πρέπει να αποδοθεί σε οργανωτικούς, πολιτισμικούς ή τεχνολογικούς παράγοντες, που θα είναι χρονιμοί στην οργάνωση κάθε παγκόσμια αναπτυσσόμενου έργου λογισμικού.

Το κύριο μέρος της ανάπτυξης του FreeBSD γίνεται στην Βόρεια Αμερική (46% των διαπράξεων). Ακολουθεί η Ευρώπη με 39% και τρίτη είναι η Ασία με 10%. Τέλος, έπονται οι Αυστραλία, Νότια Αμερική και Αφρική με 2,5%, 1,6% και 0,8% αντίστοιχα. Οι μισές σχεδόν (49%) γραμμές κώδικα αφορούσαν στον κώδικα του ίδιου του συστήματος, ενώ η μεταφορά και η δημιουργία πακέτων για υπάρκτες εφαρμογές τρίτων έφταναν το 47%. Η τεκμηρίωση και ο δικτυακός τόπος του έργου ήταν τα υπόλοιπα 2,1% και 1,7% αντίστοιχα.

Δύο ενδιαφέρουσες παραπτήρεις προκύπτουν από την κατανομή της εργασίας στις διάφορες περιοχές. Πρώτον, η δουλειά που κάνουν οι Ασιάτες διαπράξτες είναι σε πολύ μεγάλο βαθμό (80%) η μεταφορά και η δημιουργία πακέτων για υπάρχουσες εφαρμογές. Η δεύτερη παραπτήρηση αφορά σε διαφορές μεταξύ της συντήρησης και της εξέλιξης. Η περαιτέρω εξέλιξη του FreeBSD γίνεται στο τρέχοντα, ενεργό κλάδο ανάπτυξης του συστήματος διαχείρισης σχηματισμών. Φυσικά, οι προγραμματιστές ενθαρρύνονται να μεταφέρουν ("back-port") διορθώσεις και βελτιώσεις στους σταθερούς κλάδους των παλιότερων εκδόσεων του FreeBSD που συντηρούνται ακόμη μέσω μίας διαδικασίας που λέγεται "merge from current" (MFC). Οι συνεισφορές σε σταθερούς κλάδους αφορούν, κατά πάσα πιθανότητα, σε δουλειά συντήρησης. Είναι αξιοσημείωτο, όμως, ότι η κατανομή αυτών των δύο ειδών εργασίας (στο σταθερό και ενεργό κλάδο αντίστοιχα)

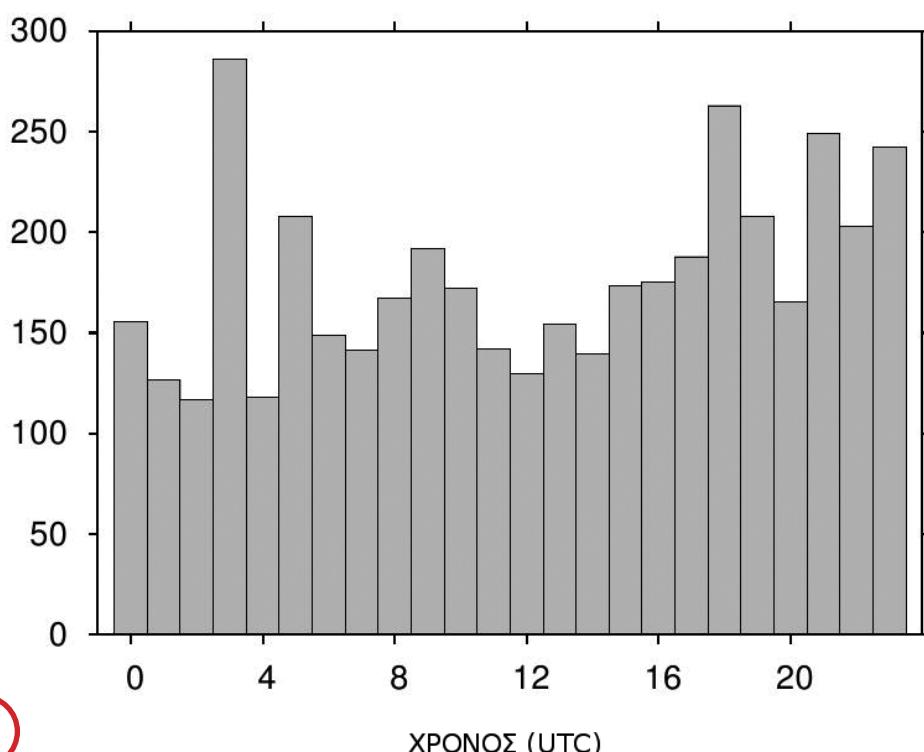
ΤΑ ΤΕΛΕΥΤΑΙΑ 10 ΧΡΟΝΙΑ, ΟΙ ΠΡΟΓΡΑΜΜΑΤΙΣΤΕΣ ΤΟΥ FREEBSD ΣΥΝΕΙΣΦΕΡΑΝ ΚΑΤΑ ΜΕΣΟ ΟΡΟ 177 ΓΡΑΜΜΕΣ ΚΩΔΙΚΑ ΑΝΑ ΩΡΑ.

Συνολικά, επιβεβαιώσαμε τα γεωγραφικά δεδομένα για 292 προγραμματιστές, χωρίς να λαμβάνουμε υπόψη όσους μετακόμισαν κατά το διάστημα που καλύπτουν τα δεδομένα μας από το CVS.

Σε διάφορες περιπτώσεις χρειάστηκε

λουμε χρόιμα συμπεράσματα.

Οι 292 προγραμματιστές του έργου για τους οποίους είχαμε γεωγραφικά δεδομένα ζουν σε 206 διαφορετικές τοποθεσίες σε όλο τον κόσμο. Οι γραμμές κώδικα που έχει συνεισφέρει κάθε τοπο-



Στο σχήμα απεικονίζεται ο μέσος αριθμός γραμμών κώδικα για κάθε ώρα της ημέρας και, όπως φαίνεται, ο FreeBSD αναπτύσσεται συνεχώς καθ' όλο το 24ωρο.

δεν είναι ίδια σε όλες τις περιοχές. Ενώ στη Βόρεια Αμερική η δουλειά περίπου ισοκατανέμεται σε συνεισφορές προς το ενεργό και τους σταθερούς κλάδους (56% και 44% αντίστοιχα), η κατανομή στην Ευρώπη και την Ασία είναι τον ενεργό κλάδο με το 70% των διαπράξεων.

Οι παραπάνω διαφορές μπορούν να εξηγηθούν με δύο θεωρίες. Η μία λέει ότι στη Βόρεια Αμερική υπάρχουν περισσότερα συστήματα παραγωγής, όπως οι διακομιστές του Yahoo! που βασίζονται στο FreeBSD, ενώ στις υπόλοιπες περιοχές του κόσμου οι περισσότεροι προγραμματιστές χρησιμοποιούν το FreeBSD στους προσωπικούς υπολογιστές τους. Στα συστήματα παραγωγής χρησιμοποιούνται συνήθως σταθερές και δοκιμασμένες εκδόσεις του λειτουργικού, έτσι μπορούμε να υποθέσουμε ότι οι προγραμματιστές στη Βόρεια Αμερική έχουν επιπλέον κίνητρα για τη συντήρηση αυτών των παλιότερων εκδόσεων. Μία εναλλακτική εξήγηση μπορεί να αφορά στη σύνθεση της ομάδας δημιουργίας των νέων εκδόσεων. Τα μέλη αυτής της ομάδας είναι υπεύθυνα για το συντονισμό των εκδόσεων του FreeBSD και ίσως έχουν μεγαλύτερο συμφέρον από τους άλλους προγραμματιστές στη συντήρηση των σταθερών εκδόσεων, είτε μόνοι τους είτε με τη βοήθεια άλλων. Ωστόσο, η σύνθεση αυτής της ομάδας δεν είναι ίδιαίτερα πολωμένη, προκειμένου να αποδοθεί εκεί η παραπάνω διαφορά. Αν και οι πέντε από τους έξι έχουν δεσμούς με τις ΗΠΑ, οι δύο έχουν επίσης δεσμούς με το Ηνωμένο Βασίλειο και ο ένας ζει μόνιμα στην Ιαπωνία.

Η γενική εικόνα περιπλέκεται και από τη διαφορετική κατανομή των προγραμματιστών που επιλύουν θέματα από τη βάση δεδομένων με τις αναφορές προβλημάτων. Εδώ η Ευρώπη με 41% έχει την πρωτοκαθεδρία, ακολουθεί η Βόρεια Αμερική με 36% των λυμένων προβλημάτων και έπονται η Ασία και η Αυστραλία με 12% και 7% αντίστοιχα. Προφανώς, η δουλειά της επιλυσης προβλημάτων διαφέρει από τη συντήρηση των σταθερών εκδόσεων.

Συνεχής ανάπτυξη

Ενα πλεονέκτημα της παγκόσμιας ανάπτυξης λογισμικού είναι η εργασία στον κώδικα σε ένα συνεχή 24ωρο κύκλο. Στο σχήμα 2, βλέπουμε ότι αυτό ακριβώς επιτυγχάνεται και στο FreeBSD. Στα τελευταία 10 χρόνια, οι προγραμματιστές του συνεισέφεραν κατά μέσο όρο 177 γραμμές κώδικα ανά ώρα κάθε μίας ημέρας. Ο αριθμός αυτός είχε ένα ελάχιστο 116 γραμμών (στις 02:00 UTC) και ένα μέγιστο των 285 γραμμών (στις 03:00 UTC).

Ενα ερώτημα σχετικό με την συνεχή 24ωρη ανάπτυξη είναι το πώς διεξάγονται οι εργασίες εξέλιξης του κώδικα. Για παράδειγμα, θα μας ενδιέφερε να δούμες αν η δουλειά σε ένα συγκεκριμένο αρχείο μεταφέρεται από μία τοποθεσία σε μία άλλη, αν συνεργάζονται οι απομακρυσμένοι γεωγραφικά προγραμματιστές σε μεγάλα δομοστοιχεία (modules) και αν οι ευθύνες ανάπτυξης κατανέμονται σε διαφορετικές περιοχές; Η απάντηση σε αυτά τα ερωτήματα μπορούν να μας οδηγήσουν σε ένα μοντέλο συνεχούς 24ωρης ανάπτυξης που θα δουλεύει σωστά. Για να δώσουμε απαντήσεις, καταγράψαμε τα ημερολόγια διαπράξεων για μεμονωμένα αρχεία, δομοστοιχεία ή ολόκληρο το σύστημα για διάφορα χρονικά διαστήματα σε 8ωρη και ημερήσια βάση. Κατόπιν, μετρήσαμε τον αριθμό των ημερών στις οποίες γίνονταν διαπράξεις, τον αριθμό των ημερών με διαπράξεις σε ένα 8ωρο από τους ίδιους ανθρώπους, τον αριθμό των ημερών που είχαν διαπράξεις την επόμενη μέρα και τον αριθμό των ημε-

δομοστοιχείο παρά τις διαφορετικές χρονικές ζώνες. Το ίδιο ισχύει και στο επίπεδο του συστήματος, όπου στο 96% των ημερών οι διαπράξεις από διαφορετικούς προγραμματιστές εκτείνονται σε 8ωρες περιόδους, ενώ μόνο στο 50% των ημερών εργασίας του έργου ο ίδιος προγραμματιστής εργάζεται για πάνω από 8 ώρες. Συμπτωματικά, η εργασία σε ένα φυσιολογικό 8ωρο της ημέρας (σε αντίθεση με την υπέρβαση του 8ώρου) φαίνεται να είναι το κυρίαρχο μοτίβο και στις τρεις κατηγορίες που εξετάσαμε. Οι διαπράξεις από τον ίδιο προγραμματιστή την επόμενη μέρα φτάνουν το 3,6% στα μεμονωμένα αρχεία, το 18% στα δομοστοιχεία και το 97,8% για όλο το σύστημα.

Παραγωγικότητα και ποιότητα

Ενα ακόμη ερώτημα που αφορά στην παγκόσμια ανάπτυξη λογισμικού είναι πώς επηρεάζει αυτό το είδος εργασίας την παραγωγικότητα των προγραμματιστών και την ποιότητα των προϊόντων

Η ΓΕΩΓΡΑΦΙΚΗ ΑΠΟΣΤΑΣΗ ΜΕΤΑΞΥ ΤΩΝ ΠΡΟΓΡΑΜΜΑΤΙΣΤΩΝ ΔΕΝ ΕΠΗΡΕΑΖΕΙ ΣΗΜΑΝΤΙΚΑ ΤΗΝ ΠΑΡΑΓΩΓΙΚΟΤΗΤΑ ΤΟΥΣ.

ρών με διαπράξεις σε 8ωρο από διαφορετικούς προγραμματιστές. Αυτοί οι αριθμοί μπορούν να μας δείξουν τόσο τα μοτίβα της "κανονικής δουλειάς" όσο και τα μοτίβα που αντιστοιχούν σε συνεχή 24ωρη ανάπτυξη.

Στο επίπεδο των μεμονωμένων αρχείων, η συνεχής ανάπτυξη δεν είναι τόσο εμφανής. Μόνο στο 1% των ημερών μία διάπραξη ακολουθούνταν από μία άλλη ενός διαφορετικού προγραμματιστή. Σε αντίθεση, οι περίοδοι εργασίας του ίδιου προγραμματιστή που ξεπερνούν τις 8 ώρες εμφανίζονται στο 1,7% των ημερών όπου γίνονταν άλλαγές σε ένα αρχείο. Προφανώς (και δικαιολογημένα) οι προγραμματιστές προτιμούν να παρατείνουν χρονικά τη διάρκεια της δουλειάς τους παρά να τη δίνουν για ολοκλήρωση σε κάποιον άλλο προγραμματιστή. Η κατάσταση αλλάζει στο επίπεδο των δομοστοιχείων. Το 5,1% των ημερών εργασίας σε ένα δομοστοιχείο περιλαμβάνουν διαπράξεις που εκτείνονται σε διάστημα μεγαλύτερο των 8 ωρών από τον ίδιο προγραμματιστή, ενώ ένα 12,5% των ημερών εργασίας περιλαμβάνουν διαπράξεις από διαφορετικούς προγραμματιστές. Εδώ φαίνεται ότι οι προγραμματιστές από διαφορετικά μέρη του κόσμου συνεργάζονται στο ίδιο

Παραγωγικότητα

Η παραγωγικότητα συνήθως μετράται ως το παραγόμενο προϊόν ανά μονάδα εισόδου. Δυστυχώς, το FreeBSD δεν κρατά δεδομένα για τις ώρες εργασίας στον κώδικα κάθε εθνολογίας. Γι' αυτό, προσπαθήσαμε να δούμες αν η γεωγραφική απόσταση μεταξύ των προγραμματιστών που δουλεύουν σε ένα δομοστοιχείο επηρεάζει τον αριθμό των γραμμών που υποβάλλουν σε αυτό. Για δάσκαλα δομοστοιχεία του FreeBSD, μετρήσαμε το μέσο όρο των γραμμών που υποβάλλονται σε ένα κυλίμενο μηνιαίο διάστημα, τη διάρκεια του χρόνου που δαπανήθηκε σε ένα συγκεκριμένο δομοστοιχείο, τον αριθμό των διαπράξεων, το πλήθος των διαφορετικών διαπρακτών και τη μέ-



ΤΟ ΣΤΥΛ ΤΟΥ FREEBSD

Οι οδηγίες για το στυλ του κώδικα περιέχονται FreeBSD Kernel Developer's Manual που είναι διαθέσιμο στο www.freebsd.org/docs.html

ση γεωγραφική απόσταση μεταξύ τους. Από τα 1.300 δομοστοιχεία που εξετάσαμε, αφαιρέσαμε όσα δεν ήταν μέρος του FreeBSD έργου (αλλά συνεισφέρονταν από τρίτους) καθώς και όσα είχαν λιγότερες από 1.000 γραμμές κώδικα διαπράξεων σε όλη την ιστορία τους. Καταλήξαμε έτσι σε 463 δομοστοιχεία.

Ως βασική υπόθεση, εξετάσαμε τη συσχέτιση μεταξύ του μέσου αριθμού των γραμμών που υποβάλλονται ανά μήνα και του ποσοστού των διαπράξεων που γίνονται από διαφορετικούς προγραμματιστές. Διαισθητικά, θα περίμενε κάποιος ότι δύο περισσότεροι άνθρωποι εργάζονται σε ένα δομοστοιχείο τόσο περισσότερο κώδικα όταν συνεισφέρουν. Πράγματι βρήκαμε μία θετική στατιστική συσχέτιση (συντελεστής συσχέτισης Pearson's: 0,67) μεταξύ των δύο μεγε-

δεδομένου ότι η αρχική υπόθεση που κάναμε φανέρωσε μία συσχέτιση μεταξύ των διαπρακτών και των γραμμών που υποβάλλουν, θα περιμέναμε μία αντίστοιχη θετική συσχέτιση μεταξύ των μεγεθών "απόσταση των διαπρακτών" και "εργασία που παράγουν". Ωστόσο, ο στατιστικός έλεγχος που κάναμε έδειξε ένα ιδιαίτερο χαμηλό ποσοστό συσχέτισης (συντελεστής συσχέτισης: -0,14) και επομένως συμπεραίνουμε ότι η γεωγραφική απόσταση μεταξύ των προγραμματιστών δεν επηρεάζει σημαντικά την παραγωγικότητά τους.

Ποιότητα κώδικα

Ενα κάποιον ενδιαφέρον ερώτημα είναι το πώς επιδρά ο μεγάλος αριθμός των (διάσπαρτων γεωγραφικά) προγραμματιστών στην ποιότητα του παραγόμενου

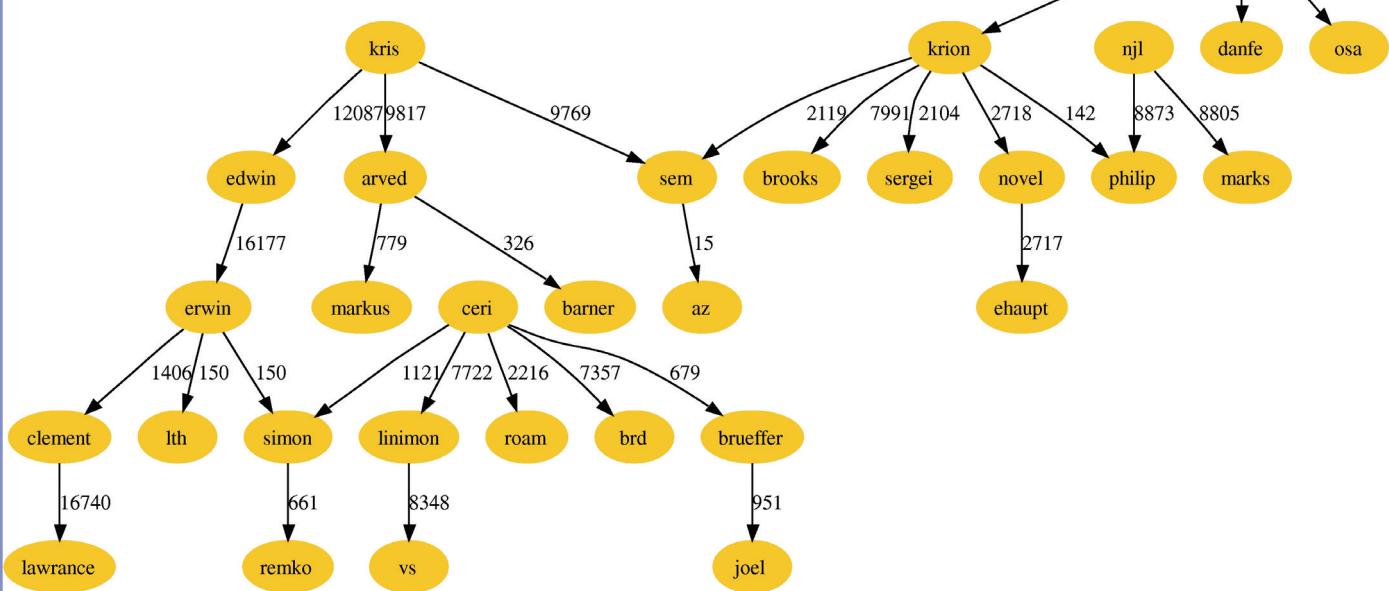
των οδηγιών για το στυλ του κώδικα του FreeBSD. Η επιλογή αυτή έγινε λόγω της ευκολίας μέτρησης της τήρησης του στυλ μέσω της μορφοποίησης κάθε αρχείου πηγαίου κώδικα με το πρόγραμμα indent, το οποίο ρυθμίσαμε με τις οδηγίες για το στυλ του FreeBSD. Με αυτό τον τρόπο, μετρήσαμε απλώς τον αριθμό των γραμμών κώδικα που θα έπρεπε να αλλάξει το indent (δηλαδή, το πλήθος των βασικών διαφορών ανάμεσα σε ένα σωστά μορφοποιημένο αρχείο και το αρχικό του). Επιπλέον, από το CVS παίραμε μία λίστα του κώδικα κάθε αρχείου, όπου κάθε γραμμή συνοδεύεται από το όνομα του ανθρώπου που την τροποποίησε τελευταίος. Ετοι μετρήσαμε και το πλήθος των προγραμματιστών που δούλεψαν σε αυτό το αρχείο.

Οταν συλλέξαμε τα δεδομένα από τα παραπάνω δύο μεγέθη, ελέγχαμε με τη μέθοδο Pearson's για τη συσχέτιση τους. Ο συντελεστής συσχέτισης για τα 11.040 ζευγάρια μετρήσεων ήταν μόλις 0,05, κάπι που σημαίνει ότι, στην περίπτωσή μας, η εμπλοκή γεωγραφικά διάσπαρτων προγραμματιστών στην ανάπτυξη του FreeBSD δεν επηρεάζει την ποιότητα του κώδικα που παράγουν.

Συχνότητα σφαλμάτων

Τέλος, μας απασχόλησε το αν η παγκόσμια ανάπτυξη ενός αρχείου κώδικα από διάφορους προγραμματιστές συνοδεύεται από έναν αυξανόμενο αριθμό αναφορών προβλημάτων για αυτό το κομμάτι κώδικα. Μία πιθανή θετική συσχέτιση θα μπορούσε να υπο-

3



Αποστάσεις σε χιλιόμετρα για τα ζευγάρια μεντόρων-μαθητευομένων του FreeBSD (φαίνεται μόνο ένα μέρος του γράφου).

δεικνύει ότι η παγκόσμια ανάπτυξη στο FreeBSD οδηγεί σε αυξανόμενο αριθμό σφαλμάτων (bugs) του κώδικα λόγω, για παράδειγμα, της κακής επικονιωνίας μεταξύ των προγραμματιστών. Αν και οι αναφορές προβλημάτων τηρούνται σε διαφορετική βάση δεδομένων από το σύστημα διαχείρισης σχηματισμών του FreeBSD, κάθε διάπραξη στο CVS που διορθώνει κάποιο πρόβλημα περιλαμβάνει συνήθως μία ετικέτα-παραπομπή στην αντίστοιχη αναφορά προβλήματος (ΑΠ). Επειδή οι αναφορές σοβαρών προβλημάτων αντιμετωπίζονται πάντα (αργά ή γρήγορα), μπορόσαμε να μετρήσουμε τη συχνότητα των ΑΠ σε ένα αρχείο, διαιρώντας τον αριθμό των διαπράξεων που έχουν ως ετικέτα τον αριθμό αυτής της ΑΠ με το συνολικό αριθμό των διαπράξεων για το συγκεκριμένο αρχείο. Μετά, εξετάσαμε τη συσχέτιση αυτού του κλάσματος με τον αριθμό των διαφορετικών προγραμματιστών που συνεισέφεραν κώδικα στο ίδιο αρχείο.

Συλλέξαμε δεδομένα για 33.392 αρχεία πηγαίου κώδικα, 457.481 διαπράξεις και 12.505 ΑΠ. Κατά μέσο όρο, κάθε αρχείο συνοδεύθηκε από 13,7 διαπράξεις, 0,27 ΑΠ και 4,2 διαφορετικούς προγραμματιστές. Ένας έλεγχος της συσχέτισης μεταξύ της συχνότητας των ΑΠ και του αριθμού των διαπράξεων μας έδειξε μία ασήμαντη συσχέτιση (συντελεστής: 0,07) και επομένως τα δεδομένα από το FreeBSD δεν υποστηρίζουν καθόλου την υπόθεση ότι η παγκόσμια ανάπτυξη λογισμικού συνοδεύεται από αυξημένη συχνότητα σφαλμάτων στον παραγόμενο κώδικα.

Ανθρώπινες απήπνευτιδράσεις

Θα ήταν όμως κοντόθαλμο να μελετήσουμε την ανάπτυξη του λογισμικού σε παγκόσμια κλίμακα μόνο όσον αφορά στο παραγόμενο προϊόν. Οι οργανισμοί εξυπηρετούν και πραγματώνουν υπαρκτές ανθρώπινες ανάγκες και κίνητρα των μελών τους, όπως είναι αυτά της απόκτησης, του δεσμάτος, της μάθησης και της υπεράσπισης. Γ' αυτό, χρησιμοποιήσαμε τα δεδομένα από το FreeBSD, προκειμένου να εξετάσουμε πώς επιδρά η απόσταση μεταξύ των προγραμματιστών στο δίκτυο των επαφών και σχέσεών τους. Συγκεκριμένα, επικεντρώσαμε την προσοχή μας σε δύο είδη σχέσεων: τη συνεργασία για την επίτευξη ενός κοι-

νού σκοπού και τη μαθητεία σε μία σχέση μέντρα-μαθητευόμενου. Για αμφότερα τα είδη σχέσεων, μετρήσαμε τη γεωγραφική απόσταση μεταξύ των σχετιζόμενων προγραμματιστών και τη συγκρίναμε με τη μέση γεωγραφική απόσταση μεταξύ όλων των προγραμματιστών (που είναι 6.701 χλμ.). Μία πολύ μικρότερη από το μέσο όρο απόσταση μεταξύ δύο συνεργαζόμενων προγραμματιστών θα σημαίνει λογικά ότι οι προγραμματιστές προτιμούν να συνεργάζονται με εκείνους που μένουν κοντά τους και από ένα τέτοιο αποτέλεσμα θα μπορούσαμε να υποθέσουμε ότι η γεωγραφική απόσταση αποτελεί εμπόδιο για τέτοιους είδους σχέσεις.

Για να δημιουργήσουμε μία λίστα με συνεργαζόμενους προγραμματιστές, σαρώσαμε τα ημερολόγια διαπράξεων σε περίσσους μίας ημέρας, ψάχνοντας για διαφορετικούς προγραμματιστές που είχαν υποβάλει κώδικα στο ίδιο αρχείο την ίδια ημέρα. Υποθέσαμε ότι τέτοιες περιπτώσεις υποδηλώνουν συνεργασία μεταξύ των αντίστοιχων προγραμματιστών, λόγω της εγγύτητας σε χρόνο και περιοχής του κώδικα. Από τα δεδομένα βρήκαμε 5.847 περιπτώσεις συνεργασίας από τις οποίες κρατήσαμε τις 4.010 για τις οποίες είχαμε παράλληλα και τις γεωγραφικές συντεταγμένες των εκάστοτε δύο προγραμματιστών. Η μέση απόσταση μεταξύ συνεργαζόμενων προγραμματιστών είναι 6.489 χλμ., δηλαδή πολύ κοντά στη μέση απόσταση μεταξύ δύο οποιονδήποτε προγραμματιστών στο FreeBSD (6.701 χλμ.). Αυτό δείχνει ότι η τεχνική συνεργασία των προγραμματιστών στο FreeBSD πολύ σπάνια επηρεάζεται από τη μεταξύ τους απόσταση.

Κατά την περίοδο που οι νέοι προγραμματιστές του FreeBSD δουλεύουν με το μέντρα τους, σημειώνουν σε όλες τις διαπράξεις τους το όνομα του μέντρα που είδε και ενέκρινε την αντίστοιχη αλλαγή. Από αυτές τις διαπράξεις μπορέσαμε να δημιουργήσουμε μία λίστα με 167 ζευγάρια μέντρα-μαθητευόμενου. Στη συνέχεια, κρατήσαμε τα 107 για τα οποία είχαμε και την τοποθεσία και των δύο μελών κάθε ζευγαριού και, χρησιμοποιώντας τις γεωγραφικές συντεταγμένες καθενός, υπολογίσαμε τη γεωγραφική απόσταση μεταξύ του μέντρα και του μαθητευόμενου (εικόνα 3). Οπως φαίνεται στον Πίνακα 1, όπου απεικονίζονται στατιστικά οι χλιομετρικές αποστάσεις μεταξύ οποιωνδήποτε δύο προγραμματι-

στών και των προγραμματιστών με σχέση μαθητείας, η μέση τιμή και η ενδιάμεση τιμή μεταξύ των μεντόρων και των μαθητών τους είναι λίγο μικρότερες από εκείνες μεταξύ δύο τυχαίων προγραμματιστών του FreeBSD. Στην περίπτωση της σχέσης μαθητείας, ακόμη και οι αριθμοί στο πρώτο τετρατμήμαριο μας δίνουν αποτάσεις μέσα σε μία μικρή χώρα ή μία πολιτεία, αλλά όχι μέσα στην ίδια πόλη.

Φαίνεται, λοιπόν, ότι κάποιες σχέσεις μαθητείας δημιουργούνται ανάμεσα σε προγραμματιστές που κατοικούν σχετικά κοντά (όπως το ζευγάρι fjoel-danfe στο πάνω δεξιό μέρος της εικόνας 3), αλλά τέτοιες σχέσεις επίσης υπάρχουν (και λειτουργούν καλά) ακόμη και όταν οι δύο προγραμματιστές ζουν σε διαφορετικές ηπείρους.

Συμπεράσματα

Από όσα είδαμε μέχρι τώρα, είναι φανερό ότι η ανάπτυξη κώδικα από μία ευρέως διασπαρμένη και με ασθενείς σχέσεις ομάδα προγραμματιστών είναι βιώσιμη πρόταση. Η παγκόσμια κατανομή των μελών της ομάδας διευκολύνει τη συνεχή ανάπτυξη καθ' όλο το 24ωρο, χωρίς πολλές αρνητικές συνέπειες στην παραγωγικότητα, στην ποιότητα του κώδικα και στη συχνότητα εμφάνισης σφαλμάτων. Επιπλέον, η συνεργασία για μία συγκεκριμένη δουλειά δεν φαίνεται να επηρεάζεται από την απόσταση. Από την άλλη, όμως, είδαμε ότι η σχέση μαθητείας σε κάποιες περιπτώσεις αναπτύσσεται πιο εύκολα μεταξύ ατόμων που κατοικούν κοντά.

Ο βαθμός στον οποίο τα συμπεράσματά μας μπορούν να χρησιμοποιηθούν στην ανάπτυξη εμπορικού λογισμικού είναι υπό συζήτηση. Η έλλειψη μίας ομάδας ελέγχου με προγραμματιστές που δουλεύουν στα ίδια γραφεία και μοιράζονται το γεύμα στην ίδια καφετέρια δεν μας επιτρέπει να επιβεβαιώσουμε το αν το FreeBSD πληρώνει ή όχι το τιμήμα της ανάπτυξης λογισμικού σε παγκόσμια κλίμακα, ακόμη και όταν δύο μέλη του μένουν στην ίδια πόλη. Από την άλλη πλευρά, υπάρχουν ήδη πολλά έργα ανάπτυξης εμπορικού λογισμικού που είναι κατανεμένα σε απομακρυσμένα γραφεία ή τοποθεσίες. Γ' αυτές τις περιπτώσεις, τα ευρήματα μας δείχνουν ότι σε περιβάλλοντα όπου οι προγραμματιστές χρησιμοποιούν καθημερινά διάφορες συνεργατικές τεχνολογίες, η απόσταση δεν αποτελεί εμπόδιο.

Εν κατακλείδι, δεδομένων των γενικά θετικών αποτελεσμάτων αυτής της έρευνας, τα έργα ανάπτυξης εμπορικού λογισμικού θα μπορούσαν, τουλάχιστον, να δοκιμάσουν να ισθετήσουν και να προσαρμόσουν κάποιες από τις πρακτικές του FreeBSD.



ΕΥΧΑΡΙΣΤΙΕΣ

Θα θέλαμε να ευχαριστήσουμε τα μέρη της κοινότητας του FreeBSD που επέτρεψαν στον υπογράφοντα να συμμετέχει στο project και παρείχαν όλα τα απαραίτητα δεδομένα και σχόλια προκειμένου να γίνει αυτή η δουλειά. Τέλος, ο Παναγιώτης Λουρίδας έκανε πολλές χρήσιμες παρατηρήσεις σε πρώιμες εκδόσεις αυτής της δουλειάς.

Πίνακας 1: Στατιστική ανάλυση των αποστάσεων μεταξύ δύο τυχαίων μελών και μεταξύ των μελών του FreeBSD που έχουν σχέση μαθητείας.

	Επάχιστη	1/4	Ενδιάμεση	Μέση	3/4	Μέγιστη
Τυχαίο ζεύγος	0	2.215	7.793	6.702	9.380	19.390
Μαθητής-Μέντορας	0	745	3.856	5.080	8.801	18.650