

Evolution of Volunteer Participation in Libre Software Projects: Evidence from Debian

Wie verhält sich die Entwicklung der freiwilligen Teilnahme in Freien oder
Open-Source-Softwareprojekten über die Zeit?
Eine quantitative Analyse am Beispiel des Debian Projektes

Timo Fleischfresser
Institut für Informatik
FU Berlin
Betreuer: Christopher Oezbek

Inhaltsverzeichnis

Einleitung	3
Ziele der Studie	3
Die Datenquellen	3
Die „sources.gz“-Dateien	4
Der Debian Popularity Contest	4
Methodik	5
Art der Datenverarbeitung	5
Die Fragestellungen	5
1) Wie viele Maintainer hat Debian und wie ändert sich die Anzahl über die Zeit?	5
2) Gibt es einen Trend in Richtung der Bildung von Maintainer Teams?	6
3) Wie viele Maintainer aus vergangenen Releases bleiben aktiv?	7
4) Was ist der Beitrag von Maintainern, die in aufeinanderfolgenden Releases aktiv bleiben?	8
5) Was passiert mit Paketen von Maintainern, die das Projekt verlassen?	8
6) Werden „wichtigere“ bzw. mehr benutzte Pakete von Maintainern mit mehr Erfahrung gepflegt?	9
Ergebnisse	10
Wie viele Maintainer hat Debian und wie ändert sich die Anzahl über die Zeit?	10
Ergebnisse der Originalstudie	10
Ergebnisse der Wiederholung	10
Zusammenfassung	13
Gibt es einen Trend in Richtung der Bildung von Maintainer Teams?	13
Ergebnisse der Originalstudie	13
Ergebnisse der Wiederholung	13
Zusammenfassung	14
Wie viele Maintainer aus vergangenen Releases bleiben aktiv?	14
Ergebnisse der Originalstudie	14
Ergebnisse der Wiederholung	15
Zusammenfassung	16
Was ist der Beitrag von Maintainern, die in aufeinanderfolgenden Releases aktiv bleiben?	17
Ergebnisse der Originalstudie	17
Ergebnisse der Wiederholung	17
Zusammenfassung	19
Was passiert mit Paketen von Maintainern, die das Projekt verlassen?	19
Ergebnisse der Originalstudie	19
Ergebnisse der Wiederholung	20
Zusammenfassung	20
Werden „wichtigere“ bzw. mehr benutzte Pakete von Maintainern mit mehr Erfahrung gepflegt?	20
Ergebnisse der Originalstudie	20
Ergebnisse der Wiederholung	21
Zusammenfassung	21
Zusammenfassung	22
Literaturverzeichnis	23

Einleitung

Ziele der Studie

„Freie“ bzw. „Open-Source-Softwareprojekte“ (*Libre Software*¹) gewinnen in der Softwareindustrie zunehmend an Bedeutung. Ob lediglich als Grundlage für weitere Entwicklungen oder als komplette Einzelanwendungen bedienen sich mehr und mehr Firmen der frei erhältlichen Softwareprodukte, die aus solchen Projekten hervorgehen.

Die Realisierung „Freier Softwareprojekte“ hängt dabei zumeist von der Arbeit freiwilliger unbezahlter Mitarbeiter (engl. Volunteers) ab, die zumeist außerhalb ihrer Arbeitszeiten Zeit investieren, um die Entwicklung der jeweiligen Anwendung voranzutreiben. Diese spezielle Form des Entwicklungsprozesses birgt einige schwer auflösbare Schwierigkeiten und Risiken, da Motivation und Engagement freiwilliger Mitarbeiter schwer vorherzusagen und zu steuern sind.

Ziel dieser Studie ist es deshalb, am Beispiel des *Debian-Projektes*² quantitative Anhaltspunkte für die Dauer und die Intensität des Engagements freiwilliger Mitarbeiter zu ermitteln. Die dabei gewonnenen Daten können dann als Anhaltspunkt für tiefergehende Studien wie beispielsweise Risikoanalysen verwendet werden.

Die Grundlage für die Arbeit bildet eine Studie von Jesus M. Gonzales-Barahona (Universität Rey Juan Carlos, Madrid) und Martin Michlmayr (Universität von Cambridge) aus dem Jahr [2005](#). Die in der Originalstudie vorgezeichneten Ansätze und Analysen wurden in der folgenden Arbeit wiederholt (teilweise mit aktuelleren Datensätzen) und an Stellen, an denen weitere Informationen sinnvoll erschienen, durch geeignete Details ergänzt. Die bearbeiteten Fragestellungen entsprechen den Fragestellungen in der Originalstudie.

Die Datenquellen

Die durchgeführten Analysen zur Entwicklung der freiwilligen Teilnehmer an „Freien Softwareprojekten“ beziehen sich im Wesentlichen auf zwei Datenquellen, die beide in den Debian-Archiven³ frei verfügbar sind und eine gute Grundlage für eine quantitative Analyse des Projekts bilden. Dies sind zum einen die „sources.gz“-Dateien, die zu jedem Release des Systems eine Beschreibung der jeweils enthaltenen Pakete liefern und zum anderen der Debian Popularity Contest, der eine Art Rangliste der am häufigsten installierten bzw. benutzten Pakete darstellt.

¹ Der Begriff „Libre Software“ bezeichnet jegliche Software, die entweder als „free software“ oder als „open source software“ lizenziert ist, also frei erhältliche Software

² Das Debian Projekt ist ein Freies Softwareprojekt, das sich mit der Entwicklung eines frei erhältlichen Betriebssystems aufbauend auf dem Linux Kernel beschäftigt. 1993 von Ian Murdock und einer Hand voll weiteren Programmierern gegründet umfasst das Projekt mittlerweile etwa 1400 freiwillige Mitarbeiter und ist damit nicht nur eines der größten Freien Softwareprojekte, sondern auch das weltweit einzige, das für jedermann frei zugänglich ist

³ <http://archive.debian.org/debian-archive/>

Die „sources.gz“-Dateien

Seit der Version 2.0 enthält das Debian Archiv zu jedem neuen Release eine Datei namens „sources.gz“, in der Informationen über die jeweils enthaltenen Pakete gespeichert sind. Dazu zählen Name des Paketes, Name des Pflegers, Version, Priorität sowie einige Zusatzinformationen. Durch die daraus resultierende Verknüpfung von Paketen und den *Maintainern*⁴, die sie pflegen, lassen sich sowohl Aussagen über die Anzahl als auch über die Arbeitsleistung der beteiligten Personen in den verschiedenen Releases treffen. Somit bilden die „sources.gz“-Dateien die Grundlage für die vorgenommenen quantitativen Analysen des Debian-Projektes. Die Studie konzentriert sich dabei nur auf die Source-Dateien der stabilen Debian Versionen seit 1998 (2.0, 2.1, 2.2, 3.0, 3.1). Die Ausnahme bildet die Version 3.1, die zum Zeitpunkt der Originalstudie noch nicht fertiggestellt war, so dass folglich noch kein stabiler Release existierte. Zum Zeitpunkt dieser Studie ist die Version 3.1 bereits fertiggestellt, so dass im Hinblick auf die Debian Version 3.1 mit aktuelleren Daten gearbeitet wird.

```
Package: 2utf
Priority: optional
Section: text
Version: 1.10
Binary: 2utf
Maintainer: Ricardas Cepas <rch@WriteMe.Com>
Architecture: any
Standards-Version: 2.3.0.1
Directory: dists/slink/main/source/text
Files:
a1b17e7e74349ff43f71b9066e3beb3f 624 2utf_1.10.dsc
16f133a9da2ab2765e30dd35f15e221c 121117 2utf_1.10.tar.gz
```

Der Debian Popularity Contest

Der *Debian Popularity Contest*⁵ ist ein Versuch, den Gebrauch der unterschiedlichen Pakete des Systems in einer Art Rangliste zu erfassen. Dazu erhält jeder Anwender die Möglichkeit ein Zusatzpaket auf seinem Rechner zu installieren. Dieses erstellt wöchentlich eine Liste der auf dem jeweiligen Rechner installierten sowie der am häufigsten verwendeten Pakete. Diese Daten werden dann per Mail an Debian gesandt, wobei der Anwender dies zuvor bestätigen muss. Bei Debian angekommene Daten werden weitestgehend anonymisiert und zu einer Gesamtangliste, dem Debian Popularity Contest, zusammengefasst. Auf der Basis dieser Rangliste lassen sich neben quantitativen Aussagen auch qualitative Aussagen über die Wichtigkeit bestimmter Pakete treffen, so dass der Debian Popularity Contest eine hilfreiche Ergänzung zu den Source-Dateien der Releases darstellt.

1	ncurses-base	9881	85	166	18	9612	(Daniel Jacobowitz)
2	apt	9880	8193	952	735	0	(Apt Development Team)
3	base-files	9880	9109	261	509	1	(Santiago Vila)
4	base-passwd	9880	8842	473	565	0	(Colin Watson)
5	bsdutils	9880	7533	1171	1176	0	(Lamont Jones)

⁴ Maintainer sind verantwortlich dafür, vorhandene Softwarepakete zu installierbaren Gesamtpaketen zusammenzufassen. Maintainer können auch die Entwickler der Pakete sein, müssen es aber nicht. In dieser Studie wird der Begriff des „Maintainers“ allerdings synonym zum Begriff des Entwicklers verwendet

⁵ <http://popcon.debian.org/>

Methodik

Art der Datenverarbeitung

Die Datenverarbeitung erfolgte semi-automatisch. Im ersten Schritt wurden die Quelldateien mittels eines Java-Programmes ausgelesen, die Datensätze entsprechend der jeweiligen Aufgabenstellung verknüpft und die Resultate sortiert in csv-Dateien (comma separated value) geschrieben, da dieses Format von Tabellenkalkulationsprogrammen gelesen werden kann. Die so erhaltenen Rohdaten mussten im zweiten Schritt von Verzerrungen durch Rechtschreibfehler, Spitznamen oder Sonderzeichen gesäubert werden, durch die Fehler beim Abzählen der Pakete pro Maintainer zustande kamen. Insgesamt wurden folgende Fehlerquellen untersucht:

- 1) Rechtschreibfehler
- 2) Spitznamen oder 2. Vornamen
- 3) Sonderzeichen
- 4) mehrere Maintainer (zu Gruppen zusammengefasst)

Im dritten und letzten Schritt erfolgte die eigentliche statistische Auswertung der gesäuberten Rohdaten mit Hilfe eines Tabellenkalkulationsprogrammes.

Die Fragestellungen

1) Wie viele Maintainer hat Debian und wie ändert sich die Anzahl über die Zeit?

Hintergrund

Hintergrund dieser Fragestellung ist die Frage, ob sich die Größe des Projektes (gemessen in Paketen) und die Anzahl der daran beteiligten Personen proportional entwickeln. Geht man davon aus, dass freiwillige, unbezahlte Mitarbeiter über die Zeit hinweg einen konstanten Beitrag für ein Projekt leisten, so müsste ein Wachstum der Projektgröße ein proportionales Wachstum der Anzahl der beteiligten Personen mit sich bringen. Untersucht wird, ob die Annahme einer konstanten Beteiligung der Maintainer am Projekt zutrifft.

Konkrete Vorgehensweise in der Originalstudie

Um die Frage nach der Entwicklung der Anzahl der Debian Maintainer im Laufe der Zeit zu beantworten, werden in den *sources.gz*-Dateien aller 5 stabilen Releases von 1998 – 2004 (Ausnahme : Version 3.1 war zum Zeitpunkt der Originalstudie noch nicht fertiggestellt => kein stabiler Release) die darin genannten Maintainer ausgezählt. Diese haben alle mindestens 1 Paket des jeweiligen Release bearbeitet, sind also aktiv an der Debian Version beteiligt. Zusätzlich werden die Gesamtanzahl der jeweils enthaltenen Pakete sowie die Gesamtanzahl der Pakete ohne „gruppenbearbeitete“ Pakete (siehe Frage 2) gezählt und der Quotient *Anzahl Pakete/Maintainer* gebildet. Dieser Quotient steht für die durchschnittliche

Arbeitsleistung eines einzelnen Maintainers des entsprechenden Release. Bei Annahme eines proportionalen Wachstums der Anzahl der Pakete und der Anzahl der Maintainer muss der Koeffizient in allen Debian Versionen stabil bleiben.

Ergänzungen

Zur Ergänzung der Frage 1 der Originalstudie enthält diese Studie nicht nur die Durchschnittswerte für die Anzahl der Pakete pro Maintainer sondern zusätzlich noch genauere Beschreibungen der entsprechenden Verteilungen. Diese werden zum einen in tabellarischer Form durch die Bildung der Standardabweichung der entsprechenden Durchschnittswerte beschrieben zum anderen in graphischer Form mit Hilfe von Boxplots. Zusätzlich erfolgt noch eine Aufspaltung der Maintainerpopulation eines jeden Release in Viel- und Wenigmaintainer und die Betrachtung der Arbeitsleistung der beiden Gruppen im Verhältnis zur Gesamtarbeit.

2) Gibt es einen Trend in Richtung der Bildung von Maintainer Teams?

Hintergrund

„Freie“ bzw. „Open-Source-Softwareprojekte“, die auf die Arbeit freiwilliger Mitarbeiter angewiesen sind, haben vor allem mit dem Problem zu kämpfen, dass freiwillige Mitarbeit schwer vorherzusagen ist. Die Tatsache, dass Maintainer jederzeit aus dem Projekt aussteigen können, birgt die Gefahr, dass das Know-how zur Wartung oder Weiterentwicklung eines von diesem Maintainer gepflegten Paketes nicht auf mehrere Personen verteilt ist, so dass Schwierigkeiten bei der Übernahme eines solchen Paketes entstehen können. Verstärkt wird dieses Problem durch die zunehmende Komplexität der einzelnen Pakete bzw. ihrer Abhängigkeiten untereinander. Eine Möglichkeit, dieser Problematik zu begegnen, ist die Gründung von Maintainer Teams, die zusammen an der Entwicklung eines einzelnen Paketes arbeiten. Dies bringt zum Einen den Vorteil, dass unterschiedliches Know-how in die Weiterentwicklung eines Paketes einfließt und zum anderen wird der Gefahr begegnet, dass der Ausstieg eines Maintainers direkt eine Verwaisung des entsprechenden Paketes zur Folge hat. Untersucht wird, ob dieses Mittel zur Eingrenzung der Probleme, die die Unvorhersehbarkeit der Beteiligung an einem „Freien Softwareprojekt“ (bzw. Open Source) mit sich bringt, im Zuge aufeinander folgender Releases auch verstärkt eingesetzt wird.

Konkrete Vorgehensweise in der Originalstudie

Zur Beantwortung der Frage nach einem Trend in Richtung der Bildung von Maintainer Teams werden in den *sources.gz*-Dateien die Pakete ausgezählt, die nicht von Einzelpersonen, sondern von Maintainer Teams bearbeitet werden. Dabei werden Pakete der Debian QA (Quality Assurance) Gruppe, einer speziellen Gruppe Debians verantwortlich für die Qualitätssicherung, getrennt betrachtet, da diese Gruppe die Pflege und Wartung von Paketen ohne Maintainer übernimmt, so dass es sich bei diesen Paketen nicht wirklich um „maintained packages“ handelt. Anschließend werden die so ermittelten Paketsummen (inklusive der Debian QA Pakete) im

Verhältnis zur Gesamtanzahl der Pakete des jeweiligen Release betrachtet und der prozentuale Anteil der gruppenbearbeiteten Pakete berechnet.

Ergänzungen

Zu Frage 2 liefert diese Studie keine methodischen Ergänzungen im Vergleich zur Originalstudie. Anzumerken ist lediglich, dass die Definition einer Gruppe in beiden Studien vermutlich unterschiedlich ist. Während die Originalstudie lediglich speziell ausgezeichnete Gruppen wie ...maintainers oder ...developers zu Gruppen zusammenfasst, werden in dieser Studie alle Pakete, die von mehr als einem Maintainer bearbeitet werden als gruppenbearbeitete Pakete deklariert.

3) Wie viele Maintainer aus vergangenen Releases bleiben aktiv?

Hintergrund

Eine grundlegende Fragestellung für das Management von Open-Source-Projekten ist die Frage nach der Dauer des Engagements der Entwickler. Da im Gegensatz zu bezahlten Tätigkeiten kein materieller Ausgleich für die geleistete Arbeit besteht, ist die Verweildauer der Entwickler in den Projekten lediglich von motivationalen Aspekten abhängig. Da die Motivation einer Person von zahlreichen Variablen abhängt und von daher schwer vorherzusagen ist, ist das Sammeln von Erfahrungswerten aus bereits bestehenden Projekten ein geeignetes Mittel, um möglicherweise Trends für die Dauer des Engagements in Open-Source-Projekten zu erkennen. Hinzu kommt, dass eine solche Größe ein gutes Vergleichskriterium zu anderen auch bezahlten Projekten darstellt.

Konkrete Vorgehensweise in der Originalstudie

Um die Anzahl der langfristig aktiven Maintainer zu bestimmen werden zunächst die Maintainer der Debian Version 2.0 ausgezählt und erfasst. Von dieser Population ausgehend wird dann für jeden neuen Debian Release geprüft, wie viele dieser Ursprungsmaintainer noch aktiv sind und wie viele Pakete sie im jeweils aktuellen Release bearbeiten. Schließlich wird aus den ermittelten Daten der sogenannte *half-life-value*⁶ gebildet, um eine Vergleichsgröße zu anderen Projekten zu gewinnen.

Ergänzungen

Als Ergänzung zur Originalstudie wird wie bereits bei Frage 1 zusätzlich zum Durchschnitt der Anzahl der Pakete pro Maintainer die Standardabweichung vom ermittelten Durchschnittswert als Maß für die Streuung der Werte berechnet.

⁶ Der half-life-value beschreibt den Zeitraum, in dem sich eine Ausgangspopulation auf die Hälfte ihrer ursprünglichen Größe reduziert hat

4) Was ist der Beitrag von Maintainern, die in aufeinanderfolgenden Releases aktiv bleiben?

Hintergrund

Neben der Frage nach der Dauer des Engagements ist auch die Frage nach dem Umfang der erbrachten Leistungen (gemessen in Paketen) der Maintainer über die verschiedenen Releases hinweg von essentieller Bedeutung für die Entwicklung eines Open-Source-Projektes. Das Ziel ist es herauszufinden, ob das Engagement von Maintainern, die in aufeinanderfolgenden Releases aktiv bleiben, steigt oder sinkt und eine prototypische Entwicklung des Umfangs der erbrachten Leistungen über die verschiedenen Releases hinweg aufzuzeigen.

Konkrete Vorgehensweise in der Originalstudie

Untersuchungsgegenstand dieser Frage sind die Maintainer, die mit einem Release neu in das Projekt einsteigen und bis zur aktuellen Debian Version 3.1 aktiv bleiben. Für jede dieser Populationen wird die Anzahl der von ihr bearbeiteten Pakete im 3.1-Release gezählt und die durchschnittliche Anzahl der Pakete pro Maintainer berechnet. Im Gegensatz zu Frage 3 erhält man somit nicht den durchschnittlichen Arbeitsaufwand einer Population über die Zeit, sondern den durchschnittlichen Arbeitsaufwand mehrerer Populationen zu einem festgelegten Zeitpunkt.

Ergänzungen

Im Vergleich zur Originalstudie wird zur Beantwortung dieser Frage zum einen zusätzlich die Standardabweichung von der durchschnittlichen Anzahl der bearbeiteten Pakete pro Maintainer berechnet, zum anderen wird die Entwicklung des Engagements der Veteranen, d.h. der über alle Releases hinweg aktiven Maintainer, im Verhältnis zum Engagement der restlichen Maintainer betrachtet. Dies geschieht vor dem Hintergrund, eine mögliche Gruppe ausfindig zu machen, die eine eventuelle Steigerung des Arbeitsaufwandes pro Maintainer über die verschiedenen Releases hinweg kompensiert.

5) Was passiert mit Paketen von Maintainern, die das Projekt verlassen?

Hintergrund

Hintergrund der Fragestellung ist die Frage, ob innerhalb von Debian ein Übernahmeprozess für verwaiste Pakete existiert. Verwaiste Pakete sind Pakete von Maintainern, die zwischen zwei aufeinanderfolgenden Releases das Projekt verlassen und ihre Pakete „unmaintained“ zurücklassen. Entsprechend der Debian Policy und den Qualitätsstandards des Projekts würden solche Pakete aus dem unmittelbar nachfolgenden Release entfernt werden, solange sie nicht von einem weiteren Maintainer übernommen werden.

Konkrete Vorgehensweise in der Originalstudie

Zur Beantwortung dieser Frage werden alle möglichen Paare von unterschiedlichen Releases betrachtet. Für jedes dieser Paare wird untersucht, welche Maintainer im jeweils älteren Release aktiv waren im jeweils aktuelleren Release aber nicht mehr aktiv sind. Genauer untersucht wird nun, ob die Pakete dieser Maintainer im neuen Release von anderen Maintainern übernommen werden oder ob sie „unmaintained“ bleiben. Dazu werden alle Pakete dieser Maintainer aus dem älteren Release gezählt und diese zunächst als „orphaned“ (verwaist) deklariert. Die Pakete, die im neueren Release dann trotzdem erscheinen, wurden von anderen Maintainern „adoptiert“. So lässt sich der Prozentsatz der adoptierten Pakete zwischen je zwei Debian Versionen berechnen, der einen Eindruck vermittelt, inwieweit ein Übernahmeprozess für verwaiste Pakete existiert. Zusätzlich wird noch der prozentuale Anteil der verwaisten Pakete im Verhältnis zur Gesamtanzahl der Pakete im alten bzw. im neuen Release ermittelt, um einen generellen Überblick zu gewinnen, ob der prozentuale Anteil verwaister Pakete steigt, sinkt oder konstant bleibt.

6) Werden „wichtigere“ bzw. mehr benutzte Pakete von Maintainern mit mehr Erfahrung gepflegt?

Hintergrund

Ein interessanter Aspekt ist die Fragestellung, ob besonders wichtige Pakete von Maintainern gepflegt werden, die schon lange im Projekt aktiv sind und somit über eine große Erfahrung in der Entwicklung des Debian Systems verfügen. Die Beantwortung dieser Frage liefert zusätzlich zu den quantitativen Aussagen der Fragen 3 und 4 qualitative Aussagen über den Inhalt der Arbeit der Maintainer. Gegenstand der Studie sind hierfür die am meisten installierten bzw. die am meisten verwendeten Pakete (siehe Debian Popularity Contest).

Konkrete Vorgehensweise in der Originalstudie

Um die Frage nach der Verteilung der wichtigen Pakete zu beantworten, muss zunächst einmal definiert werden, wie der Begriff der Wichtigkeit zu verstehen ist. In dieser Studie sind wichtige Pakete die am häufigsten installierten bzw. die am häufigsten benutzten Pakete. Auch wenn diese Pakete nicht zwangsläufig die wichtigsten Pakete nach dem allgemeinen Verständnis sein müssen, so ist doch davon auszugehen, dass Pakete, die für die Stabilität des Systems unerlässlich sind, auch mit am häufigsten installiert bzw. benutzt werden. Hinzu kommt, dass die Erfassung der am häufigsten installierten Pakete durch den Debian Popularity Contest relativ problemlos erfolgen kann, während eine Auswertung hinsichtlich der Wichtigkeit der Pakete im allgemeinen Sinn eine Evaluation aller Pakete des Systems mit sich bringen würde. Um nun eine Aussage treffen zu können, ob wichtige Pakete vor allem von erfahrenen Debian Maintainern gepflegt werden, werden die Entwickler eines jeden Release betrachtet, die ebenfalls noch in der Debian Version 3.1 aktiv sind. Gezählt werden die Installationen bzw. die „Votes“ (regelmäßig benutzte Pakete) dieser Maintainer für jeden Release entsprechend des Debian Popularity Contests. Aus der Maintainerzahl und den so ermittelten Werten werden Durchschnittswerte für die Installationen (bzw. Votes) pro Maintainer eines jeden Release gebildet. Sinken diese

Durchschnittswerte über die Zeit hinweg, so stimmt die These, dass wichtige Pakete vor allem von erfahrenen Maintainern bearbeitet werden. Andernfalls muss sie verworfen werden.

Ergebnisse

Wie viele Maintainer hat Debian und wie ändert sich die Anzahl über die Zeit?

Ergebnisse der Originalstudie

Dargestellt sind die Maintainer- und Paketzahlen für die fünf Debian Releases von 1998 (2.0) bis 2004 (unvollständige Version 3.1). Wie zu erkennen ist, steigt sowohl die Anzahl der Pakete wie auch die Anzahl der Maintainer über die Releases hinweg. Die Anzahl der Pakete steigt jedoch schneller, so dass sich mit jedem Release eine Erhöhung der durchschnittlichen Anzahl der Pakete pro Maintainer ergibt. Folglich kann die Konstanzannahme für den durchschnittlichen Aufwand der Entwickler nicht gehalten werden. Mit jedem neuen Release gibt es einen Mehraufwand für die Entwickler, der kompensiert werden muss.

Date	Rel	Mnt	Pkgs	Pkg/Mnt
Jul98	2.0	217	1101	5.1
Mar99	2.1	297	1559	5.2
Aug00	2.2	453	2601	5.7
Jul02	3.0	859	5119	6.0
Dec04	(3.1)	1237	7786	6.3

Ergebnisse der Wiederholung

Analog zur Originalstudie sind in der Tabelle die Anzahlen der Pakete und Maintainer über die Releases hinweg dargestellt. Zusätzlich enthält die Tabelle noch die Werte für die Standardabweichung vom Durchschnittswert der Pakete pro Maintainer, um einen Anhaltspunkt für den Grad der Streuung der Paketzahlen zu liefern. Die Werte bilden somit ein Qualitätskriterium für die Aussagekraft des Durchschnittswertes. Deutlicher sichtbar wird die Verteilung der Pakete in den angefertigten Boxplots (siehe unten).

Betrachtet man nun die Werte im Vergleich zur Originalstudie, so fällt auf, dass nur geringfügige Unterschiede auftreten, die auf Mappingungenauigkeiten zurückzuführen sind. Lediglich die Werte für die Debian Version 3.1 unterscheiden sich wesentlich, was daran liegt, dass diese Version zum Zeitpunkt der Originalstudie noch nicht fertiggestellt war. Nimmt man nun die Werte für die Standardabweichung vom Durchschnittswert hinzu, so werden zwei Punkte deutlich. Zum einen sind die Werte der Standardabweichung in allen Fällen größer als die entsprechenden Durchschnittswerte, was auf eine sehr große Streuung der Paketanzahlen schließen lässt, zum anderen steigt die Standardabweichung mit jedem neuen Release deutlich an. Dies könnte ein Anhaltspunkt dafür sein, dass der Mehraufwand, der durch den

unproportionalen Anstieg der Gesamtpaketzahl mit jedem neuen Release hauptsächlich von einigen wenigen Personen getragen wird.

Date	Rel	Mnt	Pkgs	Pkg/Mnt	Std.Abw.
Jul98	2.0	219	1100	5.0	5.6
Mar99	2.1	299	1557	5.2	6.4
Aug00	2.2	450	2596	5.8	7.2
Jul02	3.0	865	5113	5.9	7.8
Jun05	3.1	1308	7984	6.1	9.1

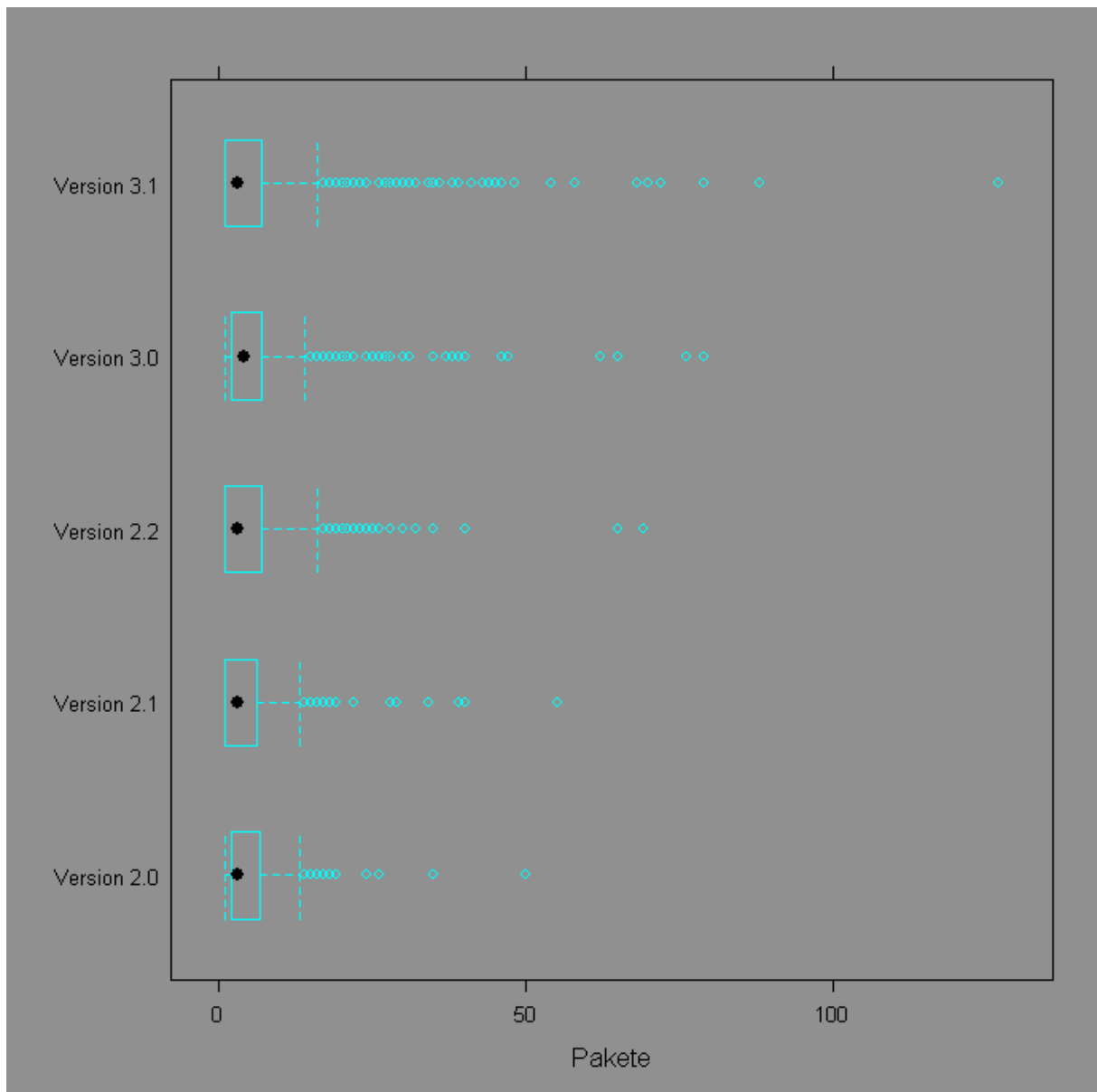
Um die Verteilung der Pakete weiter zu verdeutlichen zeigen die zwei folgenden Tabellen eine Aufschlüsselung des Engagements von Wenig- und Vielarbeitern. „Wenig“ bedeutet hier die Bearbeitung von 5 oder weniger Paketen (also unterhalb des Durchschnitts), während „viel“ die Bearbeitung von 10 oder mehr (also im oberen Bereich außerhalb der Standardabweichung) Paketen meint. Spalte 2 und 3 geben die absoluten Anzahlen der Maintainer bzw. der von ihnen bearbeiteten Pakete an, in den Spalten 4 und 5 findet man dann die entsprechenden prozentualen Anteile im Verhältnis zur Gesamtmaintainerzahl bzw. zur Gesamtanzahl der Pakete des entsprechenden Release.

Rel	Mnt 5-	Pkg 5-	Proz Mnt 5-	Proz Pkg 5-	Pkg/Mnt
2.0	156	379	71.2%	34.5%	2.4
2.1	247	471	69.9%	30.3%	1.9
2.2	301	695	66.9%	26.8%	2.3
3.0	584	1428	67.5%	27.9%	2.4
3.1	894	1990	68.3%	24.9%	2.2

Rel	Mnt 10+	Pkg 10+	Proz Mnt 10+	Proz Pkg 10+	Pkg/Mnt
2.0	33	499	15.1%	45.4%	15.1
2.1	48	787	16.1%	50.5%	16.4
2.2	81	1408	18%	54.2%	17.4
3.0	150	2747	17.3%	53.7%	18.3
3.1	233	4707	17.8%	59.0%	20.2

Dargestellt ist die Verteilung der Pakete auf die Maintainer in den verschiedenen Debian Releases in Form von Boxplots. Innerhalb des blau markierten Rechtecks befinden sich die mittleren 50% der Werte, d.h. 25% der Werte liegen links des Rechtecks und 25% der Werte liegen rechts des Rechtecks. Der schwarze Punkt markiert den Median der Verteilung also den mittleren Wert (nicht den Mittelwert). Die senkrecht verlaufenden gestrichelten Linien markieren das 10%- bzw. das 90%-Quantil.

Verteilung der Pakete auf die Maintainer in den verschiedenen Debian Versionen



Zusammenfassung

Zusammenfassend lassen sich zu Frage 1 folgende Punkte festhalten :

- 1) Die Anzahl der Maintainer steigt mit jedem Release
- 2) Die Verteilung der Pakete auf die Maintainer ist sehr ungleichmäßig
- 3) Die Unterschiede in der Verteilung werden mit jedem Release größer
- 4) Ca. 18% der Maintainer leisten über 50% der Arbeit

Gibt es einen Trend in Richtung der Bildung von Maintainer Teams?

Ergebnisse der Originalstudie

Dargestellt sind die absoluten Anzahlen der Pakete, die von Gruppen bearbeitet werden, in den verschiedenen Debian Versionen sowie der prozentuale Anteil dieser Pakete an der Gesamtpaketzahl. Die Pakete der Debian QA-Gruppe sind gesondert aufgeführt, da diese Gruppe für Pakete ohne Maintainer verantwortlich ist, also innerhalb Debians einen Sonderstatus besitzt.

Date	Release	Packages	Packages QA	Percentage
Jul98	2.0	14	14	1.3%
Mar99	2.1	21	11	1.4%
Aug00	2.2	46	31	1.8%
Jul02	3.0	101	71	2.2%
Dec04	(3.1)	599	194	7.4%

Ergebnisse der Wiederholung

Die Ergebnisse dieser Studie weichen nur leicht von den Ergebnissen der Originalstudie ab. Kleine Abweichungen kommen durch mögliche Unterschiede in der Definition von Gruppen zustande. Unterschiede in der Debian Version 3.1 sind eine Folge dessen, dass diese Version zum Untersuchungszeitpunkt der Originalstudie noch nicht fertiggestellt war.

Date	Release	Packages	Packages QA	Percentage
Jul98	2.0	15	14	1.3%
Mar99	2.1	23	15	1.5%
Aug00	2.2	51	31	1.9%
Jul02	3.0	105	71	2.0%
Jun05	3.1	745	221	8.5%

Zusammenfassung

Zusammenfassend lassen sich zu Frage 2 folgende Punkte festhalten :

- 1) Die Anzahl der von Gruppen bearbeiteten Pakete steigt mit jedem Release
- 2) Der prozentuale Anteil der von Gruppen bearbeiteten Pakete steigt mit jedem Release
- 3) Es existiert ein Trend in Richtung der Bildung von Maintainer Teams

Wie viele Maintainer aus vergangenen Releases bleiben aktiv?

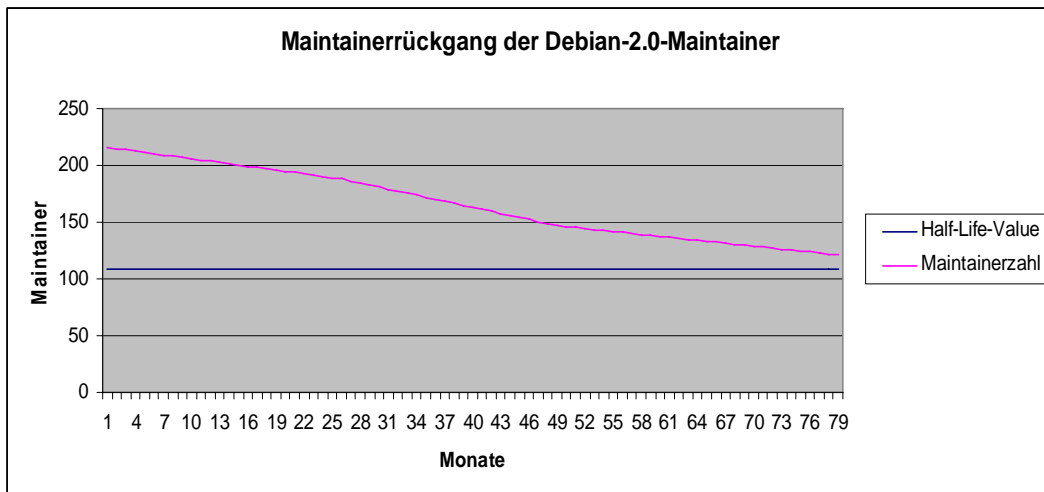
Ergebnisse der Originalstudie

Dargestellt sind die Anzahlen der in den unterschiedlichen Releases noch aktiven Debian-2.0-Maintainer, die von ihnen darin bearbeiteten Pakete sowie die durchschnittlichen Paketzahlen pro Maintainer. Betrachtet man die Werte, so wird deutlich, dass die Größe dieser Population mit jeder neuen Debian Version abnimmt. Bewertet man die ermittelten Zahlen auf Release-Basis, so hat es den Anschein, als ob die Zahl derjenigen 2.0-Maintainer, die das Projekt zwischen zwei Releases verlassen, mit jeder neuen Version steigt, wobei ein Höhepunkt zwischen den Debian Versionen 2.2 und 3.0 zu verzeichnen ist. Interpretiert man die Ergebnisse dagegen auf Zeit-Basis und interpoliert die Werte zwischen den Releasezeitpunkten, so ergibt sich ein annähernd konstanter Rückgang von ca. 1,2 Maintainern pro Monat. Der Trend, dass durchschnittlich die meisten Maintainer das Projekt zwischen den Debian Versionen 2.2 und 3.0 verlassen, bleibt jedoch bestehen (ca. 1,8/Monat). Die wenigsten Maintainer verlassen das Projekt demzufolge zwischen dem Release 3.0 und dem zur Zeit der Originalstudie noch unvollständigen Debian Release 3.1 (ca.0,9/Monat). Dies könnte damit zusammenhängen, dass zwischen den Debian Versionen 2.2 und 3.0 eventuell größere Änderungen am System vorgenommen wurden, die einen erhöhten Arbeitsaufwand für die Maintainer zur Folge hatten. Diejenigen Maintainer, die jedoch in dieser Phase des Projektes dabei bleiben, verlassen es auch innerhalb der nächsten Entwicklungsphase nicht.

Analog zur Anzahl der Maintainer sinkt auch die absolute Anzahl der von Debian-2.0-Maintainern bearbeiteten Pakete, nicht jedoch die durchschnittliche Anzahl der Pakete pro Maintainer. Die durchschnittliche Anzahl der Pakete/Maintainer steigt mit jedem neuen Release, so dass festzuhalten bleibt, dass der Rückgang der Ausgangspopulation unter anderem auch von den verbleibenden 2.0-Maintainern in Form einer Erhöhung der Paketanzahl kompensiert werden muss. Die Werte liegen allerdings unterhalb der Durchschnittswerte, die in Frage 1 für die Gesamtpopulationen der jeweiligen Debian Versionen ermittelt wurden, so dass der Anstieg der durchschnittlichen Anzahl der Pakete/verbleibenden Debian-2.0-Maintainern auch als Folge des generellen Trends zu einem Mehraufwand für den einzelnen Maintainer erklärt werden kann.

Schließlich bleibt festzuhalten, dass nach 79 Monaten zwischen dem Releasezeitpunkten der Debian Version 2.0 und der 3.1-Vorläuferversion der *half-life-value*, d.h. der Zeitraum, in der sich die Ausgangspopulation halbiert, noch nicht erreicht ist.

Date	Release	Devs	Packages	Pkg/Dev
Jul98	2.0	216	1101	5.1
Mar99	2.1	207	1086	5.2
Aug00	2.2	188	1040	5.5
Jul02	3.0	147	870	5.9
Dec04	(3.1)	121	729	6.0

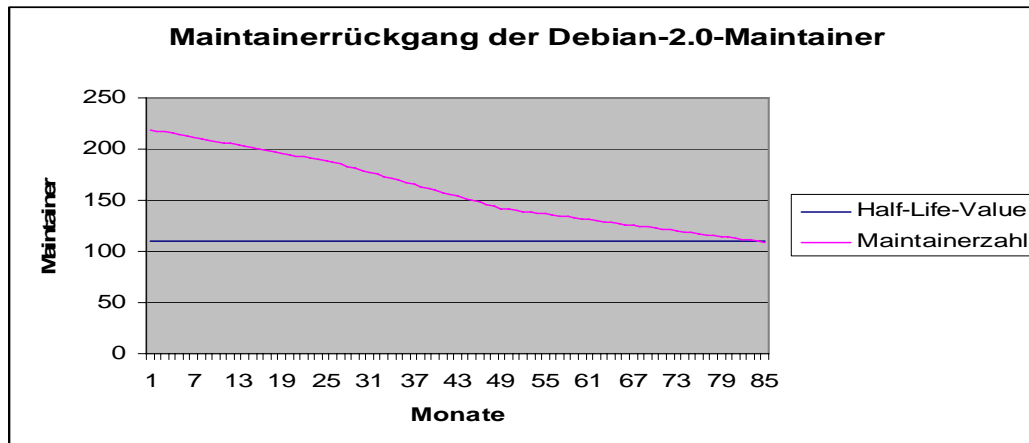


Ergebnisse der Wiederholung

Die Ergebnisse dieser Studie weichen sehr stark von den Ergebnissen der Originalstudie ab. Während die Anzahlen der verbleibenden Debian-2.0-Maintainer eines jeden Release den Werten der Originalstudie ähneln, weichen die Paketzahlen der Maintainer sehr stark von den ursprünglichen Werten ab. Als Ursachen dafür kommen entweder Mappingungenauigkeiten, die bei der manuellen Säuberung der extrahierten csv-Dateien entstehen können, oder Fehler im Auslesescript in dieser oder der Originalstudie in Frage.

Die Trends, die in der Originalstudie aufgezeigt wurden bleiben jedoch bestehen. Betrachtet man den Rückgang der 2.0-Maintainer über die Zeit, so ergibt sich auch hier ein annähernd konstanter Rückgang von ca. 1,3 Maintainern/Monat. Der stärkste Rückgang erfolgt erneut zwischen den Debian Releases 2.2 und 3.0, während zwischen den Versionen 3.0 und 3.1 durchschnittlich die wenigsten Maintainer das Projekt verlassen. Unterschiede bestehen zum einen darin, dass der *half-life-value* in dieser Studie nach etwa 84 Monaten bereits überschritten ist, zum anderen steigt die durchschnittliche Anzahl der Pakete/Maintainer für die verbleibenden 2.0-Entwickler wesentlich stärker an (auf bis zu 9,5 Pakete/Maintainer) als in der Originalstudie.

Date	Release	Devs	Packages	Pkg/Dev (St.Ab)
Jul98	2.0	219	1100	5.0 (5.7)
Mar99	2.1	209	1348	6.4 (7.2)
Aug00	2.2	187	1458	7.8 (9.2)
Jul02	3.0	142	1302	9.2 (10.7)
Jun05	3.1	109	1037	9.5 (11.9)



Zusammenfassung

Zusammenfassend lassen sich zu Frage 3 folgende Punkte festhalten :

- 1) Die Anzahl der Debian-2.0-Maintainer sinkt mit jedem Release
- 2) Interpoliert man die Release-Informationen über die Zeit, so ergibt sich durchschnittlich ein Rückgang von ca. 1,2 - 1,3 Maintainern/Monat
- 3) Zwischen den Debian Versionen 2.2 und 3.0 ist der verzeichnete Maintainerückgang am größten, zwischen den Versionen 3.0 und 3.1 ist er am geringsten
- 4) Der *half-life-value* liegt bei etwa 7 Jahren
- 5) Die durchschnittliche Anzahl der Pakete/2.0-Maintainer steigt mit aufeinander folgenden Releases
- 6) Die Werte der Originalstudie und dieser Studie für die durchschnittliche Paketanzahl/2.0-Maintainer variieren stark

Was ist der Beitrag von Maintainern, die in aufeinanderfolgenden Releases aktiv bleiben?

Ergebnisse der Originalstudie

Dargestellt sind die Maintainerzahlen der Entwickler, die in dem der Zeile entsprechenden Release erstmalig aktiv waren und auch in der Debian Version 3.1 noch aktiv sind. Zusätzlich sind in der Tabelle die Anzahlen der bearbeiteten Pakete der jeweiligen Maintainerpopulation **im Debian 3.1-Release** sowie die durchschnittlichen Paketzahlen/Maintainer abgebildet. Zu erkennen ist, dass sich die Durchschnittswerte für die Zahl der Pakete/Maintainer der verschiedenen Maintainerpopulationen nur schwach unterscheiden mit Ausnahme derjenigen Maintainer, die mit der Debian Version 2.2 zum Projekt hinzugestoßen sind. Diese scheinen besonders engagiert im weiteren Verlauf des Projektes zu sein. Darüber hinaus bleibt festzuhalten, dass sich die Zahl der Neueinsteiger (die nach wie vor aktiv sind) mit der Debian Version 3.0 sehr stark erhöht, nämlich mehr als verdoppelt hat.

Date	Release	Devs	Packages	Pkg/Dev
Jul98	2.0	121	727	6.0
Mar99	2.1	55	338	6.1
Aug00	2.2	114	919	8.1
Jul02	3.0	393	2544	6.5
Dec04	(3.1)	554	3258	5.8

Ergebnisse der Wiederholung

Die Ergebnisse dieser Studie zu Frage 4 weichen erneut sehr stark von den Ergebnissen der Originalstudie ab. Ähnlich wie bei den Resultaten der Frage 3 ähneln sich die ermittelten Maintainerzahlen (mit Ausnahme der Version 3.1), jedoch unterscheiden sich die Paketzahlen sehr deutlich, die für die jeweiligen Maintainerpopulationen ermittelt wurden. Während in der Originalstudie noch die mit der Debian Version 2.2 hinzugekommenen Maintainer die durchschnittlich engagiertesten Entwickler im aktuellen Release waren, sind es in dieser Studie die 53 Maintainer, die erstmalig im Debian Release 2.1 auftraten. Auch der ansonsten konstante Verlauf der durchschnittlichen Paketanzahl/Maintainer findet in den Ergebnissen dieser Studie keine Entsprechung. Stattdessen ist ein Rückgang des durchschnittlichen Engagements von über 9 Paketen/Maintainer der Maintainer der Debian Version 2.0 auf lediglich knapp 4 Pakete/Maintainer der Entwickler, die mit dem aktuellen Release in das Projekt eingestiegen sind, zu verzeichnen. Zum einen liegen diese Unterschiede in den aktuelleren Zahlen für die Paketzahlen der Debian Version 3.1 begründet, zum anderen erklärt dies jedoch nicht die Verdopplung der Anzahl der Pakete der Debian 2.1-Maintainer von 338 in der Originalstudie auf über 600 in dieser Studie. Hier müssen in einer der beiden Studien Fehler vorliegen.

Date	Release	Devs	Packages	Pkg/Dev (St.Ab)
Jul98	2.0	109	1037	9.5 (11.8)
Mar99	2.1	53	602	11.4 (15.6)
Aug00	2.2	106	1048	9.9 (11.5)
Jul02	3.0	389	2734	7.0 (9.3)
Jun05	3.1	651	2563	3.9 (6.4)

Um die Verteilung der Pakete noch einmal genauer zu beleuchten wurden in dieser Studie zusätzlich die absoluten Paketanzahlen, der Paketschnitt/Maintainer und der jeweilige Arbeitsanteil in Prozent der 109 Maintainer, die von der Debian Version 2.0 an dabei waren, sowie der jeweils restlichen Maintainer eines Release getrennt betrachtet. Hintergrund dabei war die Vermutung, dass der Mehraufwand in immer größer werdenden Releases sich nicht gleichmäßig auf die Maintainer verteilt, sondern hauptsächlich von bestimmten Gruppen kompensiert wird.

Wie zu erkennen ist, bewegt sich die durchschnittliche Anzahl der Pakete/Maintainer der 109 „Maintainerveteranen“ in jedem Debian Release deutlich über den entsprechenden Durchschnittswerten für die Gesamtmaintainerpopulation der Releases (vgl. Tabelle Frage 1). Mit jeder neuen Debian Version steigt die Anzahl der von ihnen bearbeiteten Pakete sowie analog dazu der jeweilige Durchschnittswert auf einen Spitzenwert von ca. 10,3 Paketen/Maintainer in der Debian Version 3.0. Erst mit dem aktuellen Debian Release 3.1 sinkt der Durchschnittswert der Veteranen leicht, obwohl die Gesamtanzahl der Pakete von der Version 3.0 zur Version 3.1 noch einmal deutlich steigt. Dies lässt die Vermutung aufkommen, dass sie ihr Engagement nach ca. 7 Jahren allmählich zurückschrauben wollen.

Im Gegensatz dazu liegen die Durchschnittswerte für die Paketanzahlen/Maintainer der restlichen Maintainer in allen Releases unterhalb des Durchschnittswertes der Gesamtmaintainerpopulation der jeweiligen Version. Zwar steigen die Werte im Zuge der immer größer werdenden Releases kontinuierlich an, überschreiten allerdings nicht das durchschnittliche Niveau der Debian Version 2.2. Dies lässt den Schluss zu, dass sich der Mehraufwand, der mit jedem neuen Release entsteht, tatsächlich nicht gleichmäßig auf die Maintainer verteilt. Ein Großteil der Arbeit wird von den Maintainerveteranen geleistet.

Release	Pkg. der 109 Vet.	Pkg/Mnt	Anteil in Prozent
2.0	686	6.3 (6.7)	61.5%
2.1	880	8.1 (8.2)	55.7%
2.2	1057	9.7 (10.3)	39.9%
3.0	1121	10.3 (11.0)	21.5%
3.1	1037	9.5 (11.9)	11.9%

Release	Pkg. Rest d. Mnt	Pkg/Mnt	Anteil in Prozent
2.0	414	3.8	37.1%
2.1	677	3.6	42.8%
2.2	1539	4.5	58.1%
3.0	3992	5.3	76.5%
3.1	6947	5.8	79.6%

Zusammenfassung

Zusammenfassend lassen sich zu Frage 4 folgende Punkte festhalten :

- 1) In der Ausgangsstudie bleibt der Beitrag von Maintainer, die in aufeinanderfolgenden Releases aktiv bleiben, relativ konstant (Ausnahme Release 2.2)
- 2) In dieser Studie sind "ältere" Maintainer wesentlich aktiver, allgemein steigt die Aktivität der Langzeitmaintainer auf durchschnittlich 10,3 Pakete/Maintainer und sinkt dann erstmalig im aktuellen Release (Rückzug ?)
- 3) Berücksichtigt man die Arbeit der 109 Veteranen, so steigt der Paketschnitt der restlichen Maintainer nicht über das Niveau der Version 2.2

Was passiert mit Paketen von Maintainer, die das Projekt verlassen?

Ergebnisse der Originalstudie

In Spalte 3 der Tabelle sind die Anzahlen der Pakete derjenigen Maintainer dargestellt, die das Projekt zwischen den zwei angegebenen Releases verlassen. Ihre Pakete sind damit zunächst verwaist (orphaned). Die Frage ist nun, wie viele dieser Pakete von anderen Maintainer übernommen (adopted) werden und im folgenden Release erneut erscheinen. Dies ist in Spalte 4 der Tabelle angegeben. Zusätzlich werden noch die „Adopted-to-Orphaned-Ratio“ (A/O), die den prozentualen Anteil der adoptierten Pakete beschreibt, sowie der prozentuale Anteil der verwaisten Pakete an der Gesamtzahl der Pakete des alten (O/Tot1) bzw. des neuen (O/Tot2) Release berechnet. Interessant sind vor allem die Zeilen, in denen die Werte zu direkt aufeinander folgenden Releases stehen. Hier werden vor allem zwei Trends sichtbar. Zum einen steigt der prozentuale Anteil der verwaisten Pakete mit jedem Release, zum anderen liegt die Wahrscheinlichkeit dafür, dass diese Pakete trotzdem übernommen werden und damit im nächsten Release erscheinen bei konstant deutlich über 60%. Dies bedeutet, es existiert innerhalb von Debian ein Übernahmeprozess, der den Fortbestand verwaister Pakete sichert. Dies bedeutet aber auch, dass mit jedem neuen Release unter der Annahme eines relativ konstanten Übernahmeprozentsatzes die Anzahl der verworfenen also nicht adoptierten Pakete steigt.

Release 1	Release 2	Orphaned	Adopted	A/O	O/Tot1	O/Tot2
2.0	2.1	15	14	93.3%	1.3%	1.0%
2.0	2.2	61	40	65.6%	5.5%	1.5%
2.0	3.0	231	171	74.0%	21.0%	4.5%
2.0	(3.1)	372	251	67.5%	33.8%	3.2%
2.1	2.2	47	31	66.0%	3.0%	1.8%
2.1	3.0	302	220	72.8%	19.4%	5.9%
2.1	(3.1)	493	327	66.3%	31.6%	6.3%
2.2	3.0	281	207	73.7%	10.8%	5.5%
2.2	(3.1)	617	403	65.3%	23.7%	7.9%
3.0	(3.1)	596	383	64.3%	11.6%	7.6%

Ergebnisse der Wiederholung

Die Ergebnisse dieser Studie unterscheiden sich nicht wesentlich von den Ergebnissen der Originalstudie. Der einzige Unterschied besteht darin, dass die ermittelten Werte der „Adopted-to-Orphaned-Ratio“ durchschnittlich etwa 5-10% über den Werten der Originalstudie liegen. Dies bedeutet, dass ein stärkerer Trend zur Übernahme verwaister Pakete festgestellt wurde als dies in der Originalstudie der Fall war.

Release 1	Release 2	Orphaned	Adopted	A/O	O/Tot1	O/Tot2
2.0	2.1	17	15	88.2%	1.5%	1.1%
2.0	2.2	67	52	77.6%	6.0%	2.5%
2.0	3.0	232	180	77.6%	20.8%	4.4%
2.0	3.1	392	273	69.6%	35.2%	4.5%
2.1	2.2	46	35	76.1%	2.9%	1.7%
2.1	3.0	292	225	77.0%	18.5%	5.6%
2.1	3.1	518	353	68.1%	32.8%	5.9%
2.2	3.0	272	215	79.0%	10.3%	5.2%
2.2	3.1	684	468	68.4%	25.8%	7.8%
3.0	3.1	727	536	73.7%	13.9%	8.3%

Zusammenfassung

Zusammenfassend lassen sich zu Frage 5 folgende Punkte festhalten :

- 1) Die Prozentzahl der verwaisten Pakete steigt mit jedem Release
- 2) Aber: konstanter Prozentsatz von >60% (ursprüngliche Studie) bzw. >70% (diese Studie) der verwaisten Pakete werden übernommen
- 3) Trotzdem steigt der prozentuale Anteil der verworfenen Pakete pro Release

Werden „wichtigere“ bzw. mehr benutzte Pakete von Maintainern mit mehr Erfahrung gepflegt?

Ergebnisse der Originalstudie

Dargestellt sind die Anzahlen der Maintainer eines jeden Release, die im aktuellen Release noch aktiv sind, sowie die Anzahlen der Installationen bzw. der Votes (regelmäßig benutzte Pakete) der von ihnen bearbeiteten Pakete entsprechend des Debian Popularity Contests. Wie zu erkennen ist, sinken sowohl der Durchschnittswert für die Anzahl der Installationen/Maintainer wie auch für die Anzahl der Votes/Maintainer über die Zeit. Dies bedeutet zunächst, dass Maintainer, die bereits länger im Projekt aktiv sind, Pakete pflegen, die durchschnittlich häufiger installiert und auch regelmäßiger benutzt werden. Die Frage, die allerdings nicht geklärt werden kann ist, ob die Implikation lauten muss „**Maintainer hat mehr Erfahrung => Maintainer pflegt Pakete, die häufiger installiert und benutzt werden**“ oder „**Maintainer ist länger im Projekt aktiv => die von ihm gepflegten Pakete existieren schon länger => die Pakete sind bekannter**“.

Release	CMaint	CPkg	Inst	Votes	Inst/Mnt	Votes/Mnt
2.0	121	729	919,856	362,249	7602.1	2993.8
2.1	176	1,066	1,306,067	498,061	7420.8	2829.9
2.2	290	1,984	2,135,137	805,642	7362.5	2778.0
3.0	683	4,528	3,712,435	1,280,173	5435.9	1874.3
(3.1)	1237	7,786	4,566,601	1,487,246	3691.7	1202.3

Ergebnisse der Wiederholung

Die Ergebnisse dieser Studie unterscheiden sich insofern von den Ergebnissen der Originalstudie, als dass die zugrunde liegende Fassung des Debian Popularity Contests aktueller ist. Die aufgezeigten Trends bleiben dabei jedoch bestehen.

Release	CMaint	CPkg	Inst	Votes	Inst/Mnt	Votes/Mnt
2.0	103	1030	1,120,944	451,124	10883.0	4379.8
2.1	158	1,653	1,638,593	653,316	10370.8	4134.9
2.2	256	2,614	2,573,854	969,917	10054.1	3788.7
3.0	612	5,262	4,396,830	1,554,005	7184.4	2539.2
3.1	1241	7,787	6,261,116	2,145,748	5045.2	1729.0

Zusammenfassung

Zusammenfassend lassen sich zu Frage 6 folgende Punkte festhalten :

- 1) Die Anzahl der Installationen bzw. der Votes pro Maintainer sinkt, je "jünger" die Population der Maintainer ist
- 2) =>populäre Pakete werden von erfahrenen Maintainern gepflegt (???)

Zusammenfassung

Versucht man, die Ergebnisse der Originalstudie und der Wiederholungsstudie zu vergleichen, so findet man die meisten Aussagen in den Resultaten beider Studien wieder. Beide zeigen deutlich einen Anstieg der Maintainerzahl wie auch verstärkt der Gesamtpaketzahl über die verschiedenen Debian Versionen hinweg. Die ermittelten Zahlen unterscheiden sich in diesem Punkt nur unwesentlich mit Ausnahme der Debian Version 3.1 aus genannten Gründen. Ebenfalls nur unwesentliche Unterschiede ergeben sich bei der Beantwortung der Frage zur Bildung von Maintainer Teams. Der Trend zur verstärkten Bildung solcher Gruppen geht sowohl aus der Original- wie auch aus der Wiederholungsstudie hervor. Auch die absoluten Zahlen sind ähnlich. Kleinere Unterschiede kommen durch eine vermutlich leicht unterschiedliche Definition des Gruppenbegriffs zustande.

Dagegen finden sich in den Resultaten der Fragen 3 und 4 größere Abweichungen zwischen den beiden Studien. Sind die Maintainerzahlen der seit der Version 2.0 aktiven Maintainer zu Frage 3 noch vergleichbar – die Ergebnisse der Wiederholungsstudie besagen einen etwas schnelleren Rückgang der ursprünglichen Maintainerpopulation (half-life-value~7 Jahre, Prognose Originalstudie~7,5 Jahre) – so unterscheiden sich die Entwicklungen der Paketzahlen der entsprechenden Maintainer deutlich (5,1 auf 6,0 in der Originalstudie, 5,0 auf 9,5 in der Wiederholungsstudie). Ein ähnlicher Trend findet sich auch in den Resultaten der Frage 4 wieder. Die Anzahlen der Maintainer, die in den verschiedenen Releases zum Projekt hinzugekommen sind und noch aktiv sind, unterscheiden sich in den beiden Studien nur leicht, während die ermittelten Paketzahlen der jeweiligen Maintainer aus dem aktuellen Debian Release sehr stark abweichen. Ergaben die Zahlen der Originalstudie mit Ausnahme der Version 2.2 einen etwa konstanten Beitrag der entsprechenden Maintainerpopulationen, so belegen die Ergebnisse der Wiederholungsstudie einen sehr hohen Beitrag der Maintainer aus den Debian Versionen 2.0, 2.1 und 2.2 sowie einen sehr starken Abfall in der Beteiligung mit den Maintainern der Version 3.0 und vor allem mit denen der Version 3.1. Die genauen Ursachen für die starken Schwankungen zwischen den beiden Studien in diesem Punkt sind unklar.

Ähnliche Ergebnisse liefern die Studien wieder in der Beantwortung der letzten beiden Fragen. Zu den Resultaten zu Frage 5 ist zu sagen, dass sowohl die absolute Anzahl wie auch der prozentuale Anteil der verwaisten Pakete mit jedem neuen Debian Release steigt. Demgegenüber steht allerdings ein Prozentsatz von konstant über 60% adoptierter Pakete in der Originalstudie und sogar von über 65% in der Wiederholungsstudie. Folglich existiert also ein Übernahmeprozess für verwaiste Pakete innerhalb des Debian Projekts. Der Vergleich der Resultate zu Frage 6 kann sich lediglich auf den Vergleich der ermittelten Entwicklungstrends konzentrieren, da unterschiedliche Versionen des Debian Popularity Contests der Auswertung zugrunde liegen und die genauen Zahlen somit nicht vergleichbar sind. Hier besteht allerdings eine Übereinstimmung insofern, als dass beide Studien belegen, dass die durchschnittliche Anzahl von Installationen bzw. Votes pro Maintainer in „älteren“ Maintainerpopulationen höher ist und diese somit vermehrt die wichtigen Pakete der Studiendefinition entsprechend pflegen. Möchte man nun allgemeine Aussagen, die aus beiden Studien hervorgehen, unterschieden nach projekthinderlichen und projektfördernden Tendenzen treffen, so lassen sich im Wesentlichen folgende Aussagen zur Entwicklung der freiwilligen Teilnahme am Debian-Projekt treffen:

Positive Entwicklungen

- Die Anzahl der Maintainer steigt mit jedem Release
- Es werden verstärkt Maintainer Teams gebildet, um der wachsenden Komplexität und der Unvorhersehbarkeit der freiwilligen Teilnahme zu begegnen
- Über 70% der verwaisten Pakete werden von anderen Maintainern übernommen

Negative Entwicklungen

- Es existiert eine große Ungleichverteilung des Engagements
- Der *half-life-value* liegt ausgehend von der Maintainerpopulation der Debian Version 2.0 bei etwa 7 Jahren, d.h. nach etwa 7 Jahren halbiert sich die Ausgangspopulation
- Der Einsatz der erfahrenen Maintainer sinkt mit dem aktuellen Release (evtl. Rückzugstendenz)
- Die Anzahl der verwaisten Pakete steigt

Literaturverzeichnis

[1] M. Michlmayr, J. M. Gonzales-Barahona, “Evolution of Volunteer Participation in Libre Software Projects: Evidence from Debian”, 2005

[2] www.Debian.org/