



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

Timo Fleischfresser
Institut für Informatik
FU Berlin
16.02.2006

Gliederung

- Einführung
- Allgemeine Vorgehensweise
- Konkrete Vorgehensweise
- Ergebnisse
- Interpretation (Diskussion)

- Was habe ich gemacht ?
 - Bearbeitung eines bereits existierenden Papers zu diesem Thema
 - Wiederholung der im Paper vorgenommenen Analysen
 - Erhöhung der Präzision der Untersuchung an Stellen, an denen unsauber gearbeitet wurde
- Warum habe ich das gemacht ?
 - Scheinerwerb
 - Ziele der Studie:
 - Quantitative Analyse der Entwicklung der Teilnahme an einem „Libre Software“-Projekt am Beispiel von Debian
 - Zukunftsprognosen
 - Generalisierung ???

- Libre Software:
 - Der Begriff „Libre Software“ bezeichnet jegliche Software, die entweder als „free software“ oder als „open source software“ lizenziert ist
- Volunteer:
 - Der Begriff „Volunteer“ bezeichnet freiwillige Mitarbeiter eines Softwareprojekts, die wirtschaftlich nicht direkt von ihrem Engagement profitieren
(„Volunteers“ können trotzdem professionelle Softwareentwickler sein)
- Maintainer:
 - Maintainer sind verantwortlich dafür, vorhandene Softwarepakete zu installierbaren Gesamtpaketen zusammenzufassen (können auch die Entwickler sein)

- Debian Projekt:
 - Gegründet 1993 von Ian Murdock
 - Freies Betriebssystem
 - Baut auf dem Linux Kern auf
 - Offen
 - Mittlerweile fast 1400 Maintainer

- **Auswertung und Analyse zweier Datenquellen:**
 - 1) Die Datei „sources.gz“ der stabilen Releases von Debian seit der Version 2.0:
 - Seit Debian 2.0 existieren im Debian Archiv zu allen Releases Beschreibungsdateien („sources.gz“), in denen alle Pakete mit Namen, Maintainer, Version, Priorität...beschrieben sind.
 - Die Studie konzentriert sich auf die stabilen Debian Versionen.
 - Benutzt werden nur die „sources.gz“-Dateien des Bereiches „main“ (nicht „cont“ und „non-free“)

Package: 2utf
Priority: optional
Section: text
Version: 1.04
Binary: 2utf
Maintainer: Ricardas Cepas <rch@debian.org>
Architecture: any
Standards-Version: 2.3.0.1
Directory: dists/hamm/main/source/text
Files:
862963920e0b0bb051416339aa8cdb67 623 2utf_1.04.dsc
54867f8ce07c343cbdc113a0d502a99f 125360 2utf_1.04.tar.gz

- Auswertung und Analyse zweier Datenquellen (Fortsetzung):

- 2) Der Debian Popularity Contest:

- Debian Benutzer können zusätzlich das Paket „popcon“ bei sich installieren. Dieses Paket sendet jede Woche eine Nachricht mit einer Liste aller beim Benutzer installierten Pakete sowie deren Benutzungszeit
- Das Ergebnis ist eine Rangliste der am häufigsten installierten bzw. benutzten Pakete

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)
6	debianutils	9880	9105	132	643	0	(Clint Adams)
7	diff	9880	9064	467	347	2	(Santiago Vila)
8	dpkg	9880	9012	210	658	0	(Dpkg Developers)
9	findutils	9880	9048	267	564	1	(Andreas Metzler)

- **Semi-automatische Datenaufnahme:**
 - Java-Programm liest die gewünschten Daten aus den Quelldateien aus und schreibt sie sortiert in eine csv-Datei
 - csv-Datei wird mit Hilfe eines Tabellenkalkulationsprogramms manuell analysiert und gesäubert (Sonderzeichen, Spitznamen, zweite Vornamen...)
- **Beantwortung von 6 Fragen:**
 - 1) Wie viele Maintainer hat Debian und wie ändert sich die Anzahl über die Zeit ?
 - 2) Gibt es einen Trend in Richtung der Bildung von Maintainer Teams ?
 - 3) Wie viele Maintainer aus vergangenen Releases bleiben aktiv ?
 - 4) Was ist der Beitrag von Maintainern, die in aufeinanderfolgenden Releases aktiv bleiben ?
 - 5) Was passiert mit Paketen von Maintainern, die das Projekt verlassen ?
 - 6) Werden „wichtigere“ Pakete bzw. mehr benutzte Pakete von Maintainern mit mehr Erfahrung gepflegt ?

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

- Vorgehensweise in der Studie:
 - Auszählen der Maintainer eines jeden Release
 - Auszählen der Pakete eines jeden Release
 - Bildung des Durchschnitts Package/Maintainer
- Ergänzungen:
 - Bilden der Standardabweichung als Maß für die Verteilung der Werte
 - Getrennte Betrachtung von Maintainern mit großem und geringem Beitrag

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

- Vorgehensweise in der Studie:
 - Auszählen der Pakete, die von Gruppen gepflegt werden
 - Berechnung des prozentualen Anteils an der Gesamtanzahl der Pakete
- Ergänzungen:
 - Eventuell andere Definition von Gruppen

1) Wie viele Maintainer aus vergangenen Releases bleiben aktiv ?

- Vorgehensweise in der Studie:
 - Auszählen der 2.0-Maintainer, die an allen Releases beteiligt sind
 - Bildung des „Half-life-values“
 - Betrachtung der Veränderung des Durchschnitts Pakete/Maintainer für diese Gruppe über die Zeit
- Ergänzungen:
 - Betrachten der Standardabweichung der Paketverteilung

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

- Vorgehensweise in der Studie:
 - Berechnung der Anzahl der Pakete am aktuellen Release der Maintainer, die mit einem bestimmten Release in das Projekt eingestiegen und nach wie vor aktiv sind
 - Berechnung des Durchschnitts Pakete/Maintainer für diese Gruppe
- Ergänzungen:
 - Verhältnis des Engagements im Vergleich zur Restpopulation
 - Betrachten der Standardabweichung der Paketverteilung

- 1) Was passiert mit Paketen von Maintainern, die das Projekt verlassen ?
 - Vorgehensweise in der Studie:
 - Betrachtung der Pakete von Maintainern, die zwischen zwei bestimmten Releases das Projekt verlassen haben
 - Unterteilung dieser Pakete in Pakete, die übernommen (adopted) oder verworfen wurden
 - Berechnung des Prozentsatzes der übernommenen Pakete
 - Berechnung des Prozentsatzes der verwaisten Pakete im Verhältnis zur Gesamtanzahl der Pakete des alten und neuen Release
 - Ergänzungen:

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

- Vorgehensweise in der Studie:
 - Auszählen der Installationen und der Votes (regelmäßig benutzte Pakete) von Maintainern, die zwischen vergangenen und dem aktuellen Release aktiv bleiben
 - Bildung eines Durchschnitts Installations/Maintainer und Votes/Maintainer
- Ergänzungen:

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

- Ergebnisse der Studie

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

Rel = Release

Mnt = Anzahl der Maintainer

Pkgs = Anzahl der Packages

Pkg/Mnt = Durchschnittliche Anzahl der Pakete pro Maintainer

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

- Meine Ergebnisse

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

Rel = Release

Mnt = Anzahl der Maintainer

Pkgs = Anzahl der Packages

Pkg/Mnt = Durchschnittliche Anzahl der Pakete pro Maintainer

StdAbw. = Standardabweichung der Pakete pro Maintainer

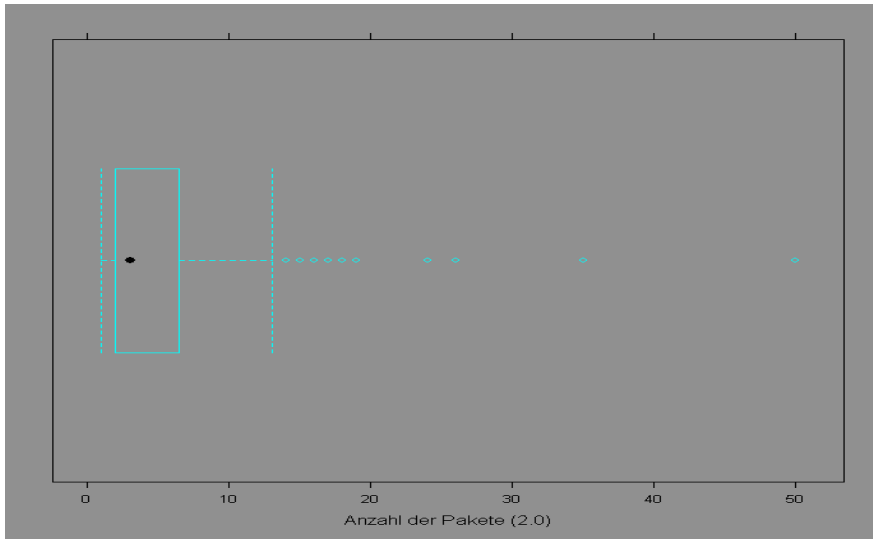
1) Wie viele Maintainer hat Debian und wie ändert sich die Anzahl über die Zeit ?

- Meine Ergebnisse

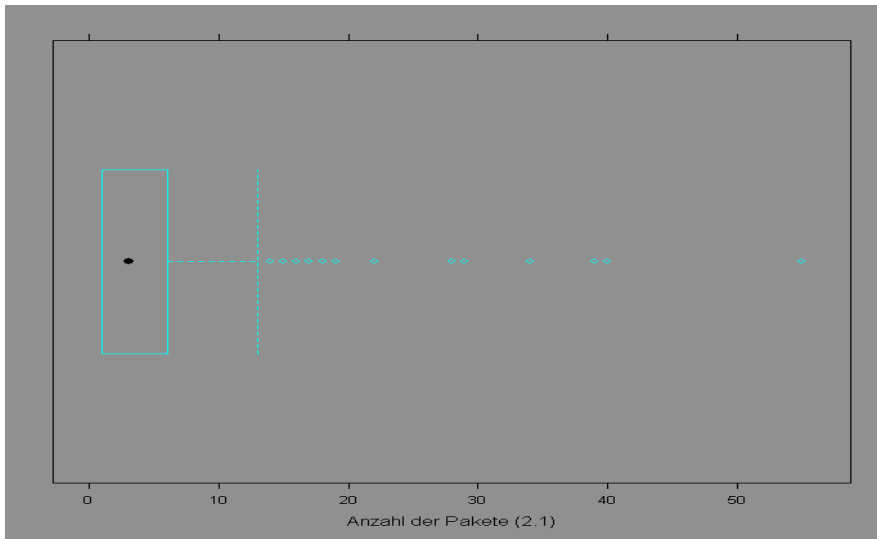
Rel	Mnt 5-	Pkg 5-	Proz Mnt 5-	Proz Pkg 5-	Pkg/Mnt
2.0	156	379	71.2%	34.5%	2.4
2.1	247	471	69.9%	30.3%	1.9
2.2	301	695	66.9%	26.8%	2.3
3.0	584	1428	67.5%	27.9%	2.4
3.1	894	1990	68.3%	24.9%	2.2

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

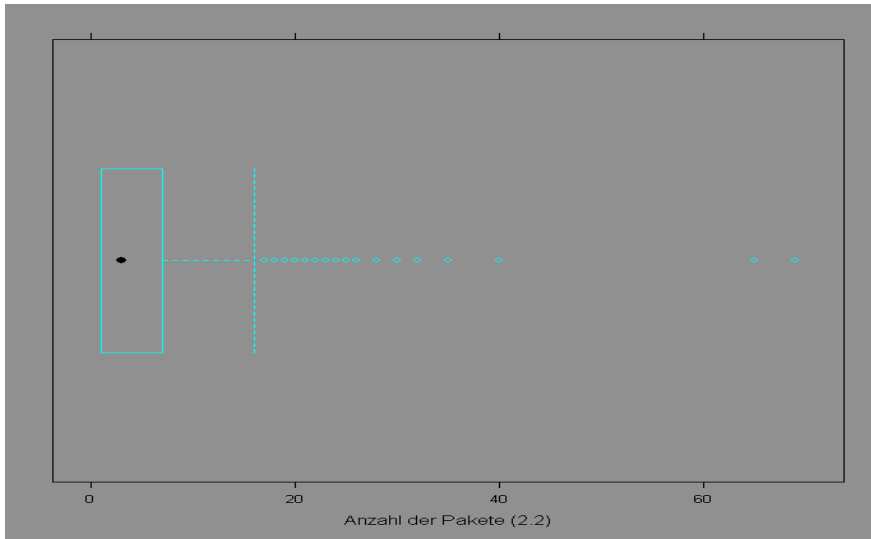
Ergebnisse



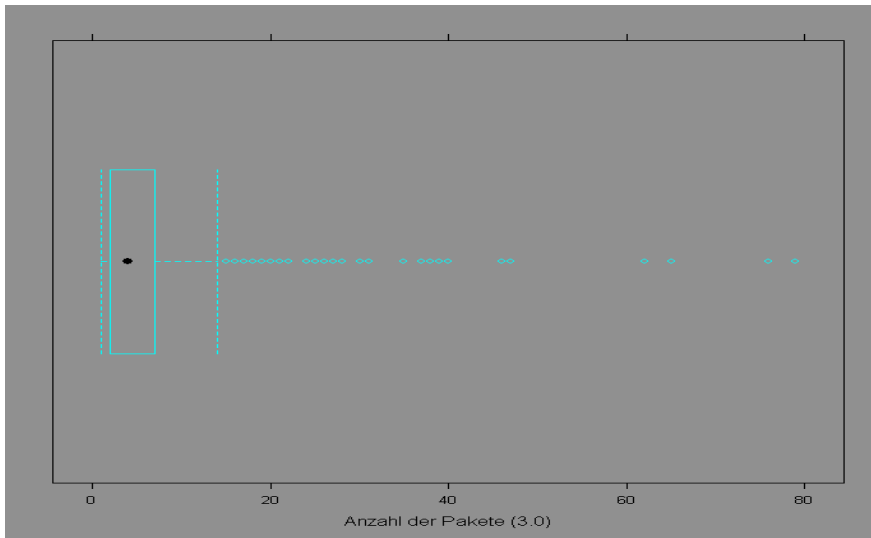
- Verteilung der Pakete auf die Maintainer in Debian 2.0



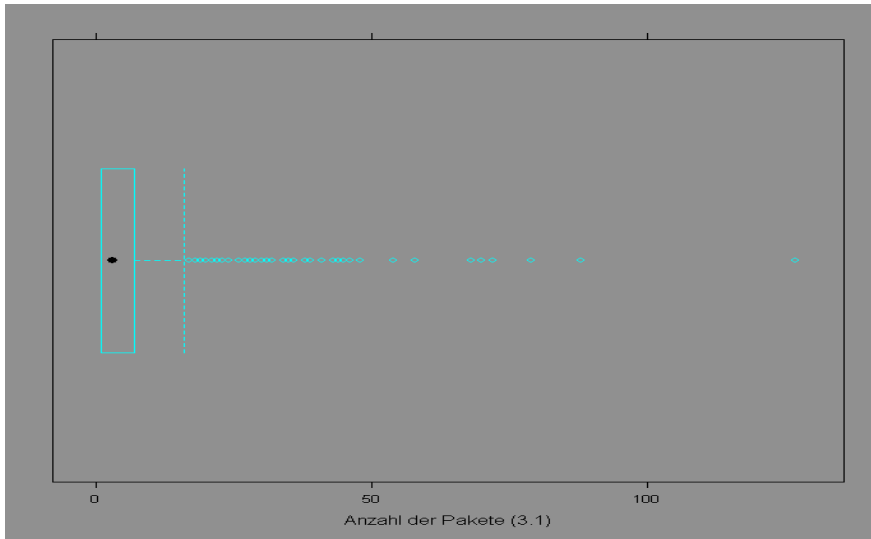
- Verteilung der Pakete auf die Maintainer in Debian 2.1



- Verteilung der Pakete auf die Maintainer in Debian 2.2



- Verteilung der Pakete auf die Maintainer in Debian 3.0



- Verteilung der Pakete auf die Maintainer in Debian 3.1

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

- Zusammenfassung
 - Anzahl der Maintainer steigt mit jedem Release
 - Anzahl der Pakete steigt mit jedem Release
 - Package/Maintainer – Rate steigt
 - Aber: Große Streuung, ca. 18% der Maintainer machen über 50% der Arbeit

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

- Ergebnisse der Studie

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%

Packages QA = Pakete der Debian QA - Gruppe

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

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%

Packages QA = Pakete der Debian QA - Gruppe

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

- Zusammenfassung
 - Anzahl der von Gruppen maintainten Pakete steigt
 - Proz. Anteil der von Gruppen maintainten Pakete steigt
 - => Trend in Richtung der Bildung von Maintainer Teams

1) Wie viele Maintainer aus vergangenen Releases bleiben aktiv ?

- Ergebnisse der Studie

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

Devs = Anzahl der Developers aus dem Release 2.0

Packages = Anzahl der Pakete der 2.0 – Developers am jeweiligen Release

1) Wie viele Maintainer aus vergangenen Releases bleiben aktiv ?

- Meine Ergebnisse

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)

Devs = Anzahl der Developers aus dem Release 2.0

Packages = Anzahl der Pakete der 2.0 – Developers am jeweiligen Release

Pkg/Dev = Durchschnitt der Anzahl der Pakete pro Developer

(in Klammern die Standardabweichung der Anzahl der Pakete pro Maintainer)

1) Wie viele Maintainer aus vergangenen Releases bleiben aktiv ?

- Zusammenfassung
 - Stetiger mehr als linearer Rückgang der Anzahl der Ursprungsmaintainer mit jedem neuen Release
 - Half-life-Value (Zeit, in der die Ausgangspopulation auf die Hälfte ihrer Größe gesunken ist) in der Ausgangsstudie noch nicht erreicht, in meiner Studie bereits überschritten (nur noch 49% der Ausgangspopulation vorhanden)
 - Durchschnittliche Anzahl der Pakete pro Maintainer steigt

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

- Ergebnisse der Studie

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

Devs = Anzahl der Developer, die im jeweiligen Release dazugekommen sind und noch aktiv sind

Packages = Anzahl der Pakete der entsprechenden Developer im aktuellen Release

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

- Meine Ergebnisse

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)

Devs = Anzahl der Developer, die im jeweiligen Release dazugekommen sind und noch aktiv sind

Packages = Anzahl der Pakete der entsprechenden Developer im aktuellen Release

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

- Meine Ergebnisse

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%

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

- Zusammenfassung
 - In der Ausgangsstudie bleibt der Beitrag von Maintainern, die in aufeinanderfolgenden Releases aktiv bleiben, relativ konstant (Ausnahme Release 2.2)
 - In meiner Studie sind "ältere" Maintainer wesentlich aktiver, allgemein sinkt die Aktivität der Langzeitmaintainer
 - Berücksichtigt man die Arbeit der 109 Veteranen, so steigt der Paketschnitt der restlichen Maintainer nicht über das Niveau der Version 2.2

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

- Ergebnisse der Studie

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%

Orphaned = Anzahl der Pakete von Developern, die das Projekt verlassen haben

Adopted = Anzahl der Pakete dieser Developer, die von anderen übernommen wurden

A/O = Prozentsatz der übernommenen Pakete

O/Tot1 = Prozentsatz der „verwaisten“ Pakete zur Paketanzahl des älteren Release

O/Tot2 = Prozentsatz der „verwaisten“ Pakete zur Paketanzahl des neueren Release

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

- Meine Ergebnisse

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%

Orphaned = Anzahl der Pakete von Developern, die das Projekt verlassen haben

Adopted = Anzahl der Pakete dieser Developer, die von anderen übernommen wurden

A/O = Prozentsatz der übernommenen Pakete

O/Tot1 = Prozentsatz der „verwaisten“ Pakete zur Paketanzahl des älteren Release

O/Tot2 = Prozentsatz der „verwaisten“ Pakete zur Paketanzahl des neueren Release

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

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

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

- Ergebnisse der Studie

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

CMaint = Maintainer aus dem jeweiligen Release, die noch aktiv sind

CPkg = Anzahl der Pakete der Maintainer am aktuellen Release

Inst = Anzahl der Gesamtinstallationen der Pakete der Maintainer

Votes = Anzahl der regelmäßig benutzten Pakete der Maintainer

Inst/Mnt = Durchschnitt der Installationen pro Maintainer

Votes/Mnt = Durchschnitt der Votes pro Maintainer

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

- Meine Ergebnisse

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

CMaint = Maintainer aus dem jeweiligen Release, die noch aktiv sind

CPkg = Anzahl der Pakete der Maintainer am aktuellen Release

Inst = Anzahl der Gesamtinstallationen der Pakete der Maintainer

Votes = Anzahl der regelmäßig benutzten Pakete der Maintainer

Inst/Mnt = Durchschnitt der Installationen pro Maintainer

Votes/Mnt = Durchschnitt der Votes pro Maintainer

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

- Zusammenfassung
 - Die Anzahl der Installationen bzw. der Votes pro Maintainer sinkt, je „jünger“ die Population der Maintainer ist
 - => populäre Pakete werden von erfahrenen Maintainern gepflegt

- Zusammenfassung der Ergebnisse:
 - Der Umfang und die Komplexität des Systems steigt !!!
 - Welche Auswirkungen hat dies ?
 - 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
 - Die Anzahl der erfahrenen Maintainer sinkt
 - Der Einsatz der erfahrenen Maintainer sinkt
 - Die Anzahl der verwaisten Pakete steigt

- Interne Validität:
 - Sehr hoch, da kein fehlerbehafteter Versuchsaufbau, sondern reale Daten
 - Höchstens Verzerrung des Popularity Contests durch freiwillige Installation des PC => möglicherweise nur bestimmte Gruppe, die das entsprechende Paket installiert
- Externe Validität:
 - ???

- Gefahren für die Generalisierbarkeit der Ergebnisse:
 - Betriebssystem ist eine allgemein benötigte Anwendung
 - Was passiert bei speziellen Anwendungen
 - Betriebssystem stellt eine hohe technische Herausforderung dar
 - Was passiert bei Anwendungen, die technisch weniger anspruchsvoll aber trotzdem arbeitsintensiv sind
 - „Innovations-Prinzip“
 - Was passiert, wenn (Konkurrenz-)Anwendungen gebaut werden sollen, für die bereits Realisierungen existieren

- www.debian.org/
- <ftp://ftp.debian.org/debian/>
- <http://archive.debian.org/debian-archive/>
- <http://popcon.debian.org/>
- Evolution of Volunteer Participation in Libre Software Projects: Evidence from Debian (Gonzales-Barahona, Michlmayr)



Danke