



**«Московский государственный технический
университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)**

ФАКУЛЬТЕТ «Информатика и системы управления»
КАФЕДРА «Защита информации»

ЛАБОРАТОРНАЯ РАБОТА

по теме:

**«Сравнение эффективности работы программ архивации/сжатия»
(по дисциплине «Системное ПО»)**

Выполнил:

Студент ИУ10-36

Можжухин Алексей Николаевич

Постановка задачи

Требуется изучить семь архиваторов (zip, compress, gzip, rar, tar, pkzip и ha), установить их на хостовую или гостевую ОС и протестировать их эффективность на семи видах файлов (pdf, dll, txt, cab, wav, html, mp4), проанализировав различия и особенности.

Методика проведения сравнения

Архиваторы tar, zip и gzip были установлены в гостевой ОС Ubuntu 22 изначально, а rar и compress (он же ncompress) быстро устанавливались стандартным способом через «sudo apt-get install». Архиватор ha пришлось искать на сторонних ресурсах, его версия под Windows не запускалась на 64-битной хостовой системе, поэтому была найдена сборка под Ubuntu на amd64. Пробная версия архиватора pkzip для Windows была найдена на официальном сайте, он представляет собой программу с графическим интерфейсом (в отличие от всех остальных архиваторов на Ubuntu).

Характеристики для хостовой ОС (Windows 10 64-bit v21H1) следующие: Intel Core i3-8130U CPU, 12 GB RAM, 256GB SSD.

Для гостевой ОС (Ubuntu 22.04.1 LTS 64-bit через Oracle VM VirtualBox): Доступны все 4 логических ядра на 100%, 4 GB RAM, 10GB VDI.

Насчет используемых для сравнения файлов стоит отметить следующее: в качестве файла Multimedia был взят видеофайл формата MP4 с битрейтом 2 мегабита в секунду, а в качестве файла Signal использовался аудиофайл формата WAV с битрейтом 128 килобит в секунду.

Результаты сравнений эффективности архиваторов

tar:

Таблица 1. Результаты, полученные с помощью программы tar

Тип файла	Размер до, кБайт	Размер после, кБайт	Степень сжатия, %	Затраченное время, с
PDF	862,2	870,4	-0,9510554396	<1
DLL	10016	10020	-0,03993610224	<1
TXT	36,4	41	-12,63736264	<1
CAB	2260,9	2270	-0,4024945818	<1
WAV	478,8	481,3	-0,52213868	<1
HTML	39,5	51,2	-29,62025316	<1
MP4	30062,4	30070	-0,02528074937	<1

zip:

Таблица 2. Результаты, полученные с помощью программы zip

Тип файла	Размер до, кБайт	Размер после, кБайт	Степень сжатия, %	Затраченное время, с
PDF	862,2	701,5	18,63836697	<1
DLL	10016	3001,3	70,03494409	<1
TXT	36,4	1,4	96,15384615	<1
CAB	2260,9	2221,6	1,738245831	<1
WAV	478,8	395,6	17,37677527	<1
HTML	39,5	3,2	91,89873418	<1
MP4	30062,4	29983,2	0,2634520198	<1

rar:

Таблица 3. Результаты, полученные с помощью программы rar

Тип файла	Размер до, кБайт	Размер после, кБайт	Степень сжатия, %	Затраченное время, с
PDF	862,2	686,3	20,401299	<1
DLL	10016	2703,2	73,01118211	<1
TXT	36,4	1,1	96,97802198	<1
CAB	2260,9	2224,5	1,609978327	<1
WAV	478,8	396,5	17,18880535	<1
HTML	39,5	2,9	92,65822785	<1
MP4	30062,4	29988,4	0,246154665	3

compress:

Таблица 4. Результаты, полученные с помощью программы compress

Тип файла	Размер до, кБайт	Размер после, кБайт	Степень сжатия, %	Затраченное время, с
PDF	862,2	-	-	-
DLL	10016	4028,4	59,78035144	<1
TXT	36,4	11	69,78021978	<1
CAB	2260,9	-	-	-
WAV	478,8	445,2	7,01754386	<1
HTML	39,5	8,9	77,46835443	<1
MP4	30062,4	-	-	-

gzip:

Таблица 5. Результаты, полученные с помощью программы gzip

Тип файла	Размер до, кБайт	Размер после, кБайт	Степень сжатия, %	Затраченное время, с
PDF	862,2	701,4	18,64996521	<1
DLL	10016	3001,1	70,03694089	<1
TXT	36,4	1,3	96,42857143	<1
CAB	2260,9	2221,5	1,742668849	<1
WAV	478,8	395,5	17,39766082	<1
HTML	39,5	3,1	92,15189873	<1
MP4	30062,4	29983,1	0,2637846612	2

pkzip:

Таблица 6. Результаты, полученные с помощью программы pkzip

Тип файла	Размер до, кБайт	Размер после, кБайт	Степень сжатия, %	Затраченное время, с
PDF	862,2	685	20,55207608	<1
DLL	10016	2992	70,12779553	<1
TXT	36,4	2	94,50549451	<1
CAB	2260,9	2223	1,676323588	<1
WAV	478,8	387	19,17293233	<1
HTML	39,5	4	89,87341772	<1
MP4	30062,4	30013	0,1643248709	1

ha:

Таблица 7. Результаты, полученные с помощью программы ha

Тип файла	Размер до, кБайт	Размер после, кБайт	Степень сжатия, %	Затраченное время, с
PDF	862,2	701,8	18,60357226	<1
DLL	10016	2942,8	70,61900958	1
TXT	36,4	1,1	96,97802198	<1
CAB	2260,9	2253,6	0,3228802689	<1
WAV	478,8	399,2	16,62489557	<1
HTML	39,5	2,8	92,91139241	<1
MP4	30062,4	30062,5	-0,0003326414391	5

Выводы по результатам проведенных сравнений

Ввиду малого размера файлов, они архивируются без ощутимых затрат по времени. Разницу во времени архивации получается увидеть на примере видеофайла (30 МБайт), причем ни один из архиваторов не сжимает видеофайл эффективно.

Архиватор tar не подразумевает сжатие, он упаковывает один или несколько файлов или директорий в один файл, поэтому работает быстро. Целесообразно использовать его для пересылки небольших объемов данных между пользователями UNIX-систем, так как для Windows это не самый распространенный формат.

Архиватор zip является оптимальным вариантом, так как достаточно эффективно сжимает большинство файлов и является одним из самых популярных форматов.

Архиватор rar незначительно превосходит zip по эффективности сжатия и почти настолько же популярен. Из рассматриваемых архиваторов именно rar лидирует по сжатию DLL библиотек, при архивации больших объемов таких библиотек эти 3 процента разницы могут стать решающими при выборе архиватора.

На фоне остальных архиваторов негативно выделился compress, оказавшись неспособным не только сжать некоторые файлы, а вообще архивировать их (не создается файл с расширением «.Z»). Причем рассматриваемая версия (ncompress) подразумевает сжатие уже упакованного .tar файла, что не добавляет удобства в использовании утилиты compress. В рамках данной работы не удалось сформулировать аргументы в пользу использования именно compress при наличии альтернатив.

Утилита `gzip` не показала заметного превосходства над `zip`. Целесообразно использовать ее на UNIX-системах. Часто встречаются архивы «.tar.gz».

Программа `pkzip` показала более высокую эффективность сжатия PDF документов и аудиофайлов по сравнению с `zip`. Имеет смысл использовать именно это решение на системах Windows с перспективой интеграции других продуктов этой компании.

Выделилась из ряда архиваторов утилита `ha`. Учитывая то, что последняя версия вышла около 11 лет назад (если смотреть `github`-репозиторий автора), эффективность сжатия некоторых видов файлов опередила свое время. Однако, по результатам сравнений скорость сжатия оказалась ниже, чем у других архиваторов.

Итого, для достижения максимального баланса между удобством использования и эффективностью сжатия стоит выбирать между `zip` и `rar`. Второй вариант немного эффективнее сжимает (например, PDF файлы) и поддерживается и на Windows, и на UNIX. Однако, `zip` популярнее и изначально встроен во многие UNIX системы, что означает немного более удобный опыт использования, по сравнению с `rar`, который нужно сначала установить.