

Оптимизация метода сжатия страниц памяти с использованием подсчета информационной энтропии

Студент: Хамзина Регина Ренатовна ИУ7-83Б
Научный руководитель: Оленев Антон Александрович

Москва, 2023

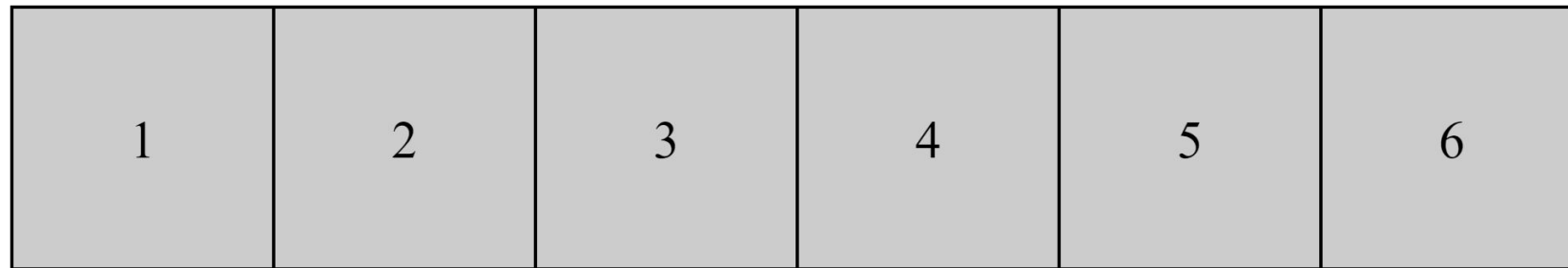
Цель и задачи

Цель: разработка оптимизации метода сжатия страниц памяти с использованием подсчета информационной энтропии.

Задачи:

- провести анализ предметной области сжатия данных оперативной памяти и вычисления информационной энтропии;
- разработать оптимизацию метода сжатия страниц памяти с использованием подсчета энтропии;
- разработать программное обеспечение, реализующее данную оптимизацию;
- провести исследования корреляции энтропии и показателей качества сжатия, соотношения времени сжатия и времени вычисления энтропии, зависимостей показателей качества и времени сжатия от типов хранимых в памяти данных.

Сжатие страниц памяти



N

Перестали использоваться страницы 2-5



x y

$x < y < N$, где N — размер страницы, x — размер сжатых данных страницы 2, y — размер сжатых данных страницы 3.

Методы подсчета энтропии

Критерий сравнения	Метод скользящего окна	Биномиальный метод
Временная сложность	$O(N + 2^n)$	$O(N + n^2)$
Базовая вычислительная арифметика	+	-
Возможность распараллеливания вычислений	+	+
Объем требуемой дополнительной памяти	2^n	n

N — размер страницы, n — длина рассматриваемой подпоследовательности битов.

Связь сжатия данных и информационной энтропии

Информационная энтропия вычисляется по формуле:

$$H(X) = - \sum_{i=0}^{255} (p_i \cdot \log_2 p_i),$$

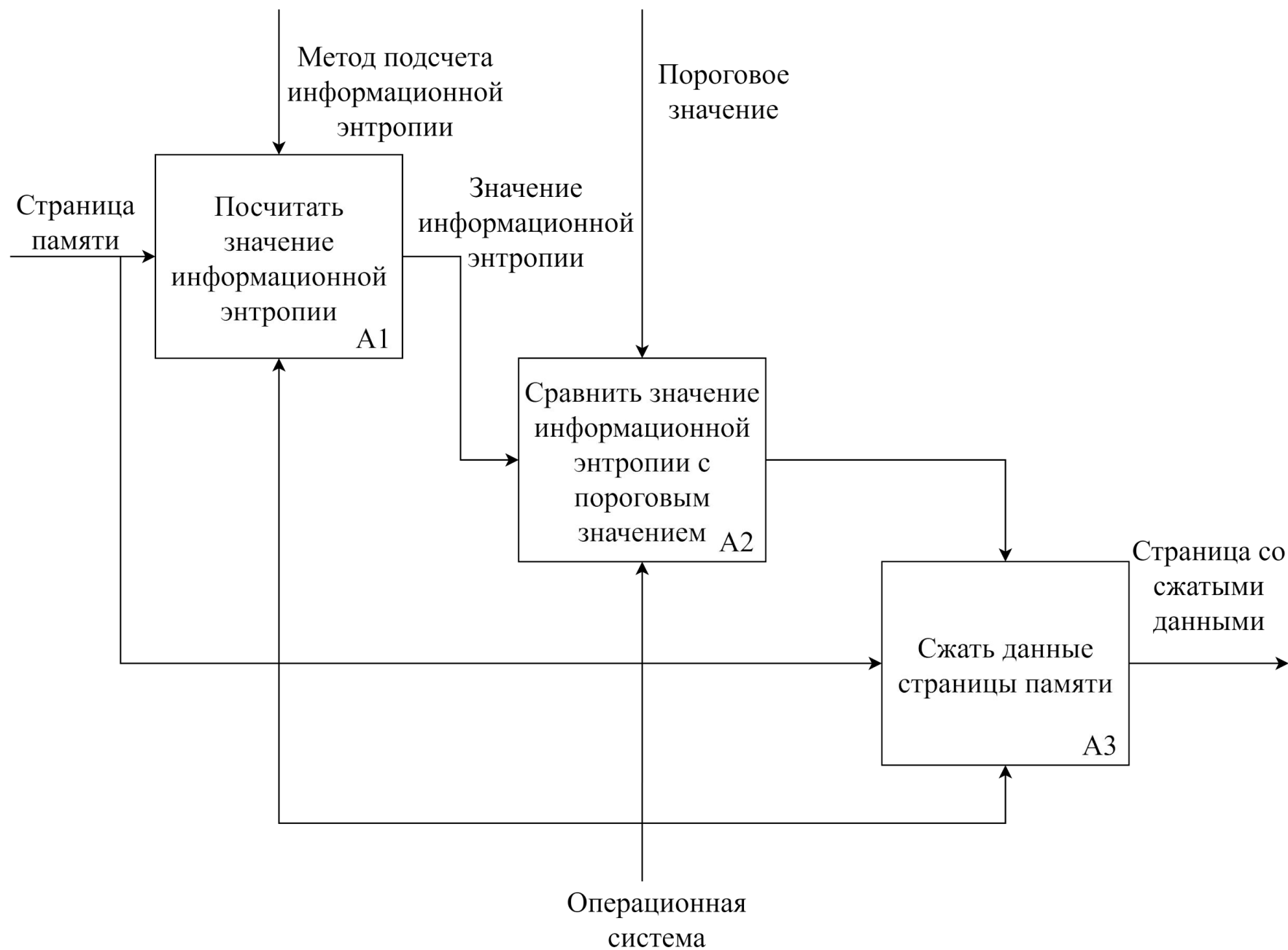
где p_i — вероятность появления байта в массиве байтов.

Коэффициент сжатия определяется по формуле:

$$K_{\text{сж}} = \frac{L_{\text{исх}}}{L_{\text{сж}}},$$

где $L_{\text{исх}}$ — объем исходных данных, $L_{\text{сж}}$ — объем сжатых данных.

Метод сжатия страниц памяти с использованием подсчета информационной энтропии



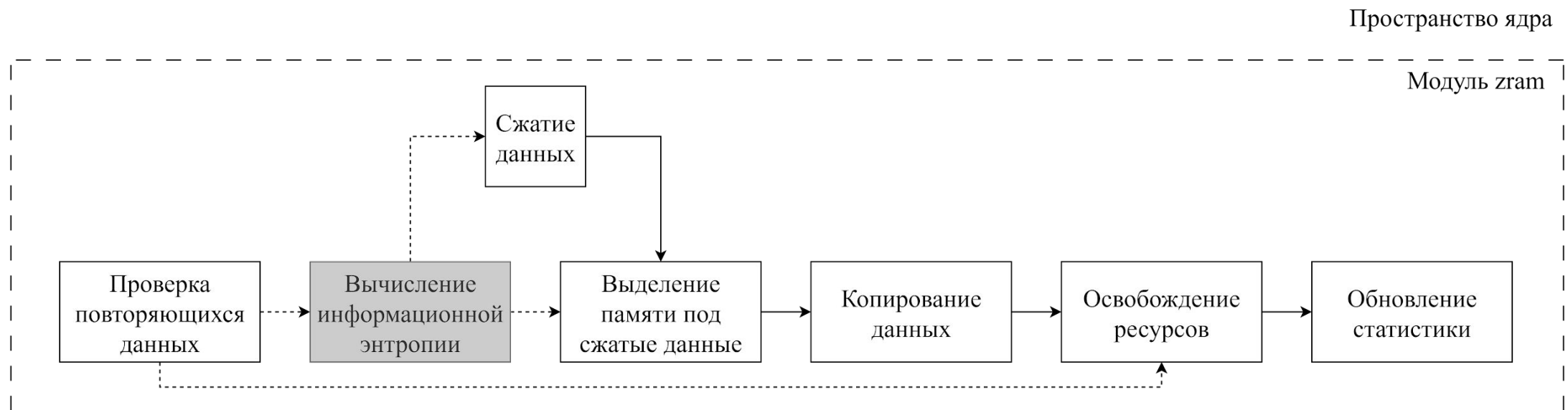
Операционные системы

Операционная система	Поддержка сжатия страниц оперативной памяти	Открытый исходный код	Доля рынка, %
Windows	+	-	63.13
macOS	+	-	17.78
Linux	+	+	2.83
FreeBSD	+	+	0.01

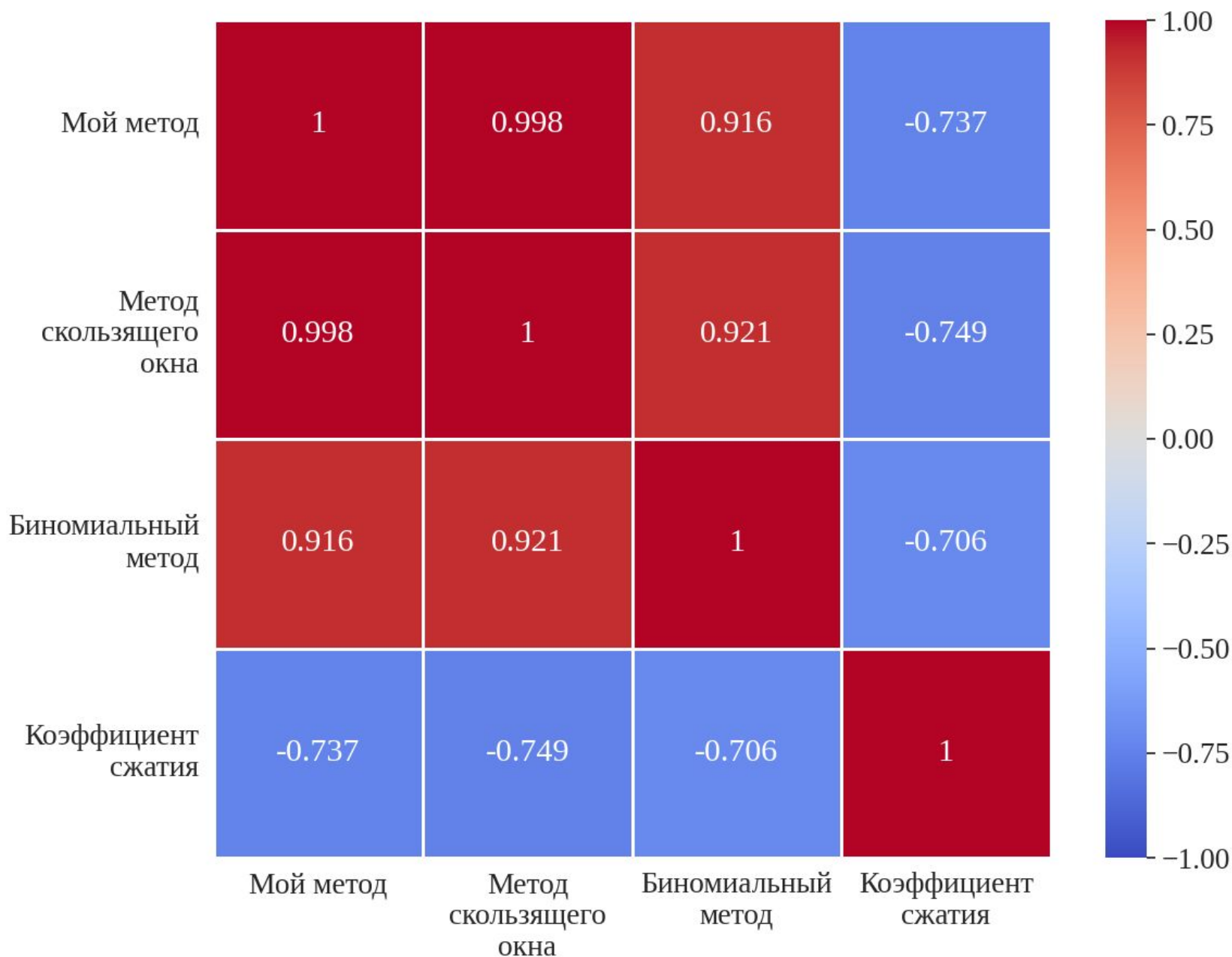
Модули ядра Linux, предоставляющие сжатие страниц оперативной памяти

Критерий сравнения	Модуль zram	Модуль zswap
Необходимость устройства подкачки	-	+
Возможность выбора алгоритма сжатия	+	+
Многопоточное сжатие	+	-
Возможность повторного сжатия с другим алгоритмом сжатия	+	-

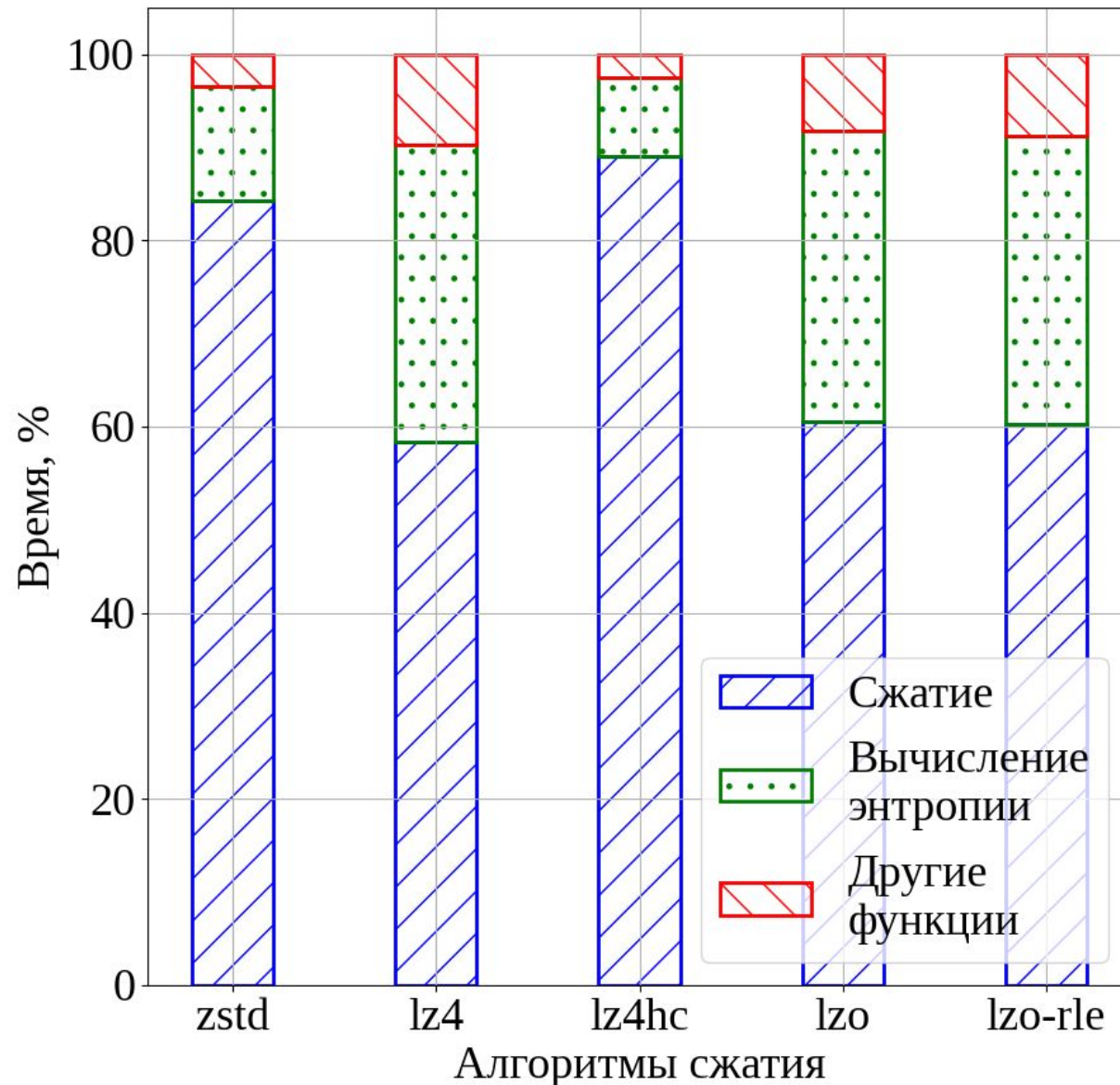
Модификация модуля zram



Корреляция энтропии и коэффициента сжатия



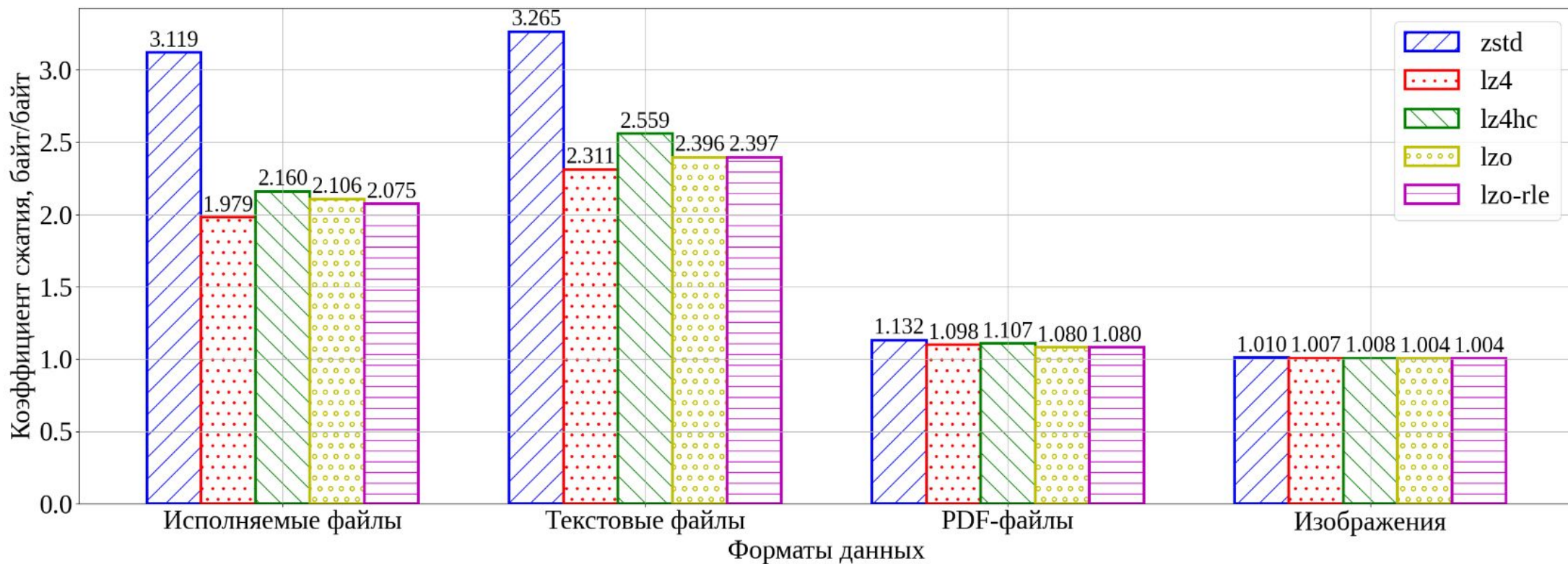
Соотношение времени сжатия и времени вычисления энтропии



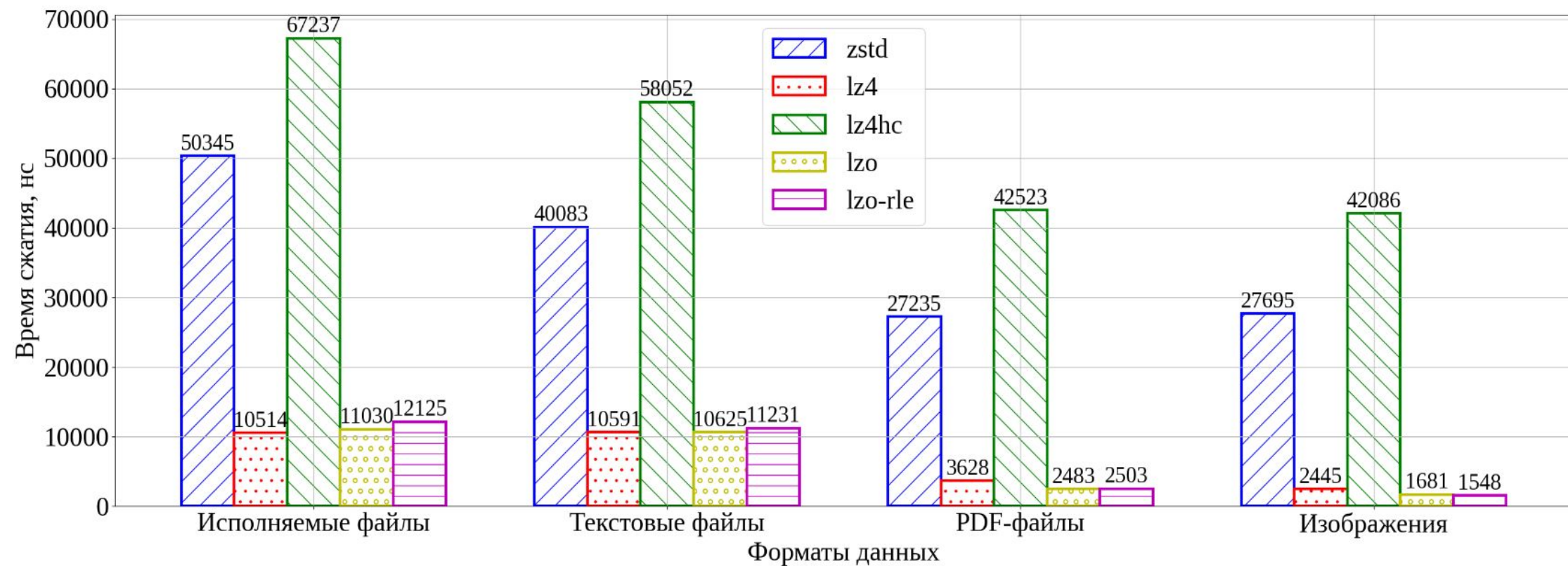
Классификация сжимаемых данных

Тип данных	Источник данных
Исполняемые файлы	Директория /usr/bin
Текстовые файлы	Исходный код ядра Linux
PDF-файлы	Открытый набор резюме livescareer.com
Изображения	Открытый набор изображений цветов

Зависимость коэффициента сжатия от типов хранимых в памяти данных



Зависимость времени сжатия от типов хранимых в памяти данных



Заключение

Была разработана оптимизация метода сжатия страниц памяти с использованием подсчета информационной энтропии.

Были выполнены следующие задачи:

- проведен анализ предметной области сжатия данных оперативной памяти и вычисления информационной энтропии;
- разработана оптимизация метода сжатия страниц памяти с использованием подсчета энтропии;
- разработано программное обеспечение, реализующее данную оптимизацию;
- проведены исследования корреляции энтропии и показателей качества сжатия, соотношения времени сжатия и вычисления энтропии, зависимостей показателей качества и времени сжатия от типов хранимых в памяти данных.

Дальнейшее развитие

- сбор и обновление статистики значений информационной энтропии страниц процессов для уменьшения времени обработки последующих страниц;
- распараллеливание вычисления информационной энтропии.