

## Программное обеспечение для одновременного захвата и записи мультигигабитного потока РСДБ-данных, передаваемого по сети Ethernet

© И. А. Безруков<sup>1</sup>, А. И. Сальников<sup>1</sup>, В. А. Яковлев<sup>1</sup>,  
А. В. Вылегжанин<sup>1,2,3</sup>

<sup>1</sup>ИПА РАН, г. Санкт-Петербург, Россия

<sup>2</sup>ФТИ им. А. Ф. Иоффе, г. Санкт-Петербург, Россия

<sup>3</sup>СПБО МСЦ РАН — филиал ФГУ ФНЦ НИИСИ РАН, г. Москва, Россия

В статье представлено обоснование выбора структуры системы буферизации и передачи данных, описано разработанное для этой системы программное обеспечение на основе специализированного средства обработки высокоскоростного потока данных netmap [1]. Подробно описаны состав аппаратного обеспечения (модель сервера и дискового массива, процессоры и сетевые устройства, используемый объем оперативной памяти), а также системы буферизации и передачи данных. Приведены результаты работы программного обеспечения в составе системы буферизации и передачи данных при проведении наблюдений за период 2016–2019 гг. на радиотелескопах РТ-13 в обсерваториях «Бадары», «Зеленчукская» и «Светлое» РСДБ-комплекса «Квазар-КВО».

**Ключевые слова:** хранение данных, регистрация данных, передача данных.

<https://doi.org/10.32876/ApplAstron.51.20-25>

### Введение

Радиоинтерферометр нового поколения, созданный в ИПА РАН на основе антенны с диаметром зеркала 13.2 м [2], предназначен для оперативно-го обеспечения космической навигационной системы ГЛОНАСС поправками всемирного времени. В настоящее время в ИПА РАН эксплуатируется система буферизации и передачи данных (СБПД) радиоинтерферометрических наблюдений.

Разработанное в ИПА РАН программное обеспечение (ПО) для СБПД осуществляет захват и запись потока данных со скоростью до 32 Гбит/с. Поток генерируется широкополосной системой преобразования сигналов (ШСПС) радиотелескопа [3]. Для регистрации 8 потоков данных от ШСПС со скоростью 2 Гбит/с в каждом канале требуется скорость записи 16 Гбит/с. Оценки возможности повышения скорости записи до 32 Гбит/с предназначены для будущих работ.

В статье представлено обоснование выбора параметров СБПД, описаны основные технологии и компоненты программного обеспечения захвата и записи данных (английское название ПО — pkt-caprec).

## Обоснование выбора программной платформы и архитектуры СБПД

В 2012 г. авторами были проанализированы основные требования к характеристикам системы: при количестве каналов до 8 скорость в канале — 2–8 Гбит/с. Были проведены многочисленные тесты различных макетов аппаратно-программных конфигураций системы. В результате проделанной работы были выработаны следующие критерии выбора конфигурации СБПД:

*a.* Необходимая аппаратная платформа — стандартное серверное оборудование, серийно изготавливаемое и имеющееся в продаже, с возможностью выбора из нескольких производителей;

*b.* Необходимость отказа от использования аппаратных RAID-контроллеров из-за их низкой производительности (для решения поставленной задачи);

*c.* Необходимая операционная система (ОС) — ОС подобная UNIX, с открытым исходным кодом, под свободными лицензиями BSD, GPL и т. д.;

*d.* Необходимая файловая система — система, входящая в стандартную поставку ОС, поддерживаемая широким сообществом разработчиков, с поддержкой логических томов в режимах stripe, mirror, raid5, raid6;

*e.* Необходимость использования отличного от стандартного, но поддерживаемого ОС, метода захвата пакетов из-за высокой скорости потока данных РСДБ;

*f.* Необходимость поддержки различных вариантов подключения ШСПС к СБПД для обеспечения гибкости и надежности системы в целом:

- один канал ШСПС → один порт СБПД;
- один канал ШСПС → коммутатор ethernet → один порт СБПД.

В 2012 г. авторам были известны две разрабатываемые системы СБПД для РСДБ-систем нового поколения: Mark 6 (обсерватория Haystack MIT [4]) и FlexBuff (Европейский объединенный институт РСДБ JIVE [5]).

Система Mark 6 не соответствовала следующим критериям: *a* — в качестве аппаратной платформы использовалось специально разработанное оборудование компанией Conduant, и *d* — в качестве файловой системы была использована специально разработанная файловая система; также система Mark 6 имела ограничение в виде регистрации только 4 каналов.

Система FlexBuff не соответствовала критериям *d* — использовалась специально разработанная надстройка над файловой системой XFS (на уровне FUSE FS) и *f* — объединялись все каналы ШСПС на одном интерфейсе, что в данном случае дало бы суммарную скорость 16 Гбит/с, которая не гарантирует успех обработки потока с одного сетевого интерфейса.

В результате анализа доступных на 2012 г. вариантов оптимальным оказалось сочетание следующих параметров:

— по критерию *b* — ОС FreeBSD (начало разработки: версии 9, 10 и в данный момент 11);

— по критерию *d* — файловая система ZFS (Zettabyte File System);

— по критерию *e* — системный API FreeBSD netmap;

— по критериям *a, f* — любая серверная платформа, поддерживаемая ОС с необходимым числом сетевых адаптеров 10 Гбит/с Ethernet.

В соответствии с этими критериями в 2012–2018 гг. на радиотелескопах РТ-13 обсерваторий РСДБ-комплекса «Квазар-КВО» была создана СБПД, описанная ниже.

## Аппаратно-программные средства

Аппаратной платформой СБПД являются серверы Dell PowerEdge R720 или R740 в конфигурации с двумя процессорами Intel Xeon (E5-2660 v2 или Gold 5120 соответственно), 256 ГБ оперативной памяти и подключенные к ним дисковые полки SuperMicro (модель SuperChassis 847E, до 44-х HDD). В каждый сервер устанавливаются четыре двухпортовые сетевые карты Intel X520 (контроллер Intek 82599), что суммарно с установленными на материнскую плату сервера составляет десять сетевых интерфейсов 10 Гбит Ethernet, восемь из которых используются для захвата данных от ШСПС, а два предназначены для передачи данных в центр корреляционной обработки (ЦКО РАН).

По состоянию на конец 2019 г. СБПД работает под управлением операционной системы FreeBSD (версия 11.1), для записи и хранения данных используется файловая система ZFS.

## Захват пакетов на сетевом интерфейсе

Для обработки высокоскоростного потока пакетов на интерфейсе 10 Гбит/с Ethernet, в ПО pkt-caprec используются API netmap-ядра ОС FreeBSD. В netmap API использованы преимущества современных контроллеров Ethernet напрямую (в данном случае Intel 82599), без сетевого стека ОС, базирующегося на классической архитектуре mbuf. При этом пользовательский код netmap (в основном это операции с пакетными буферами) выполняется на непривилегированном уровне ОС, что существенно снижает издержки, возникающие на переключении контекстов CPU между ядром ОС и приложением.

Отметим основные методы, используемые netmap, для ускорения пакетного ввода-вывода между приложением и сетью:

- группированный ввод/вывод (I/O batching);
- оптимизированные драйвера сетевых (Ethernet) контроллеров;
- предварительное выделение (аллокация) буферов ввода/вывода;
- отображение памяти буферов ввода/вывода.

Программный код netmap API включен в ядро FreeBSD с 2011 г. и поддерживается сообществом разработчиков ОС.

Структурная схема работы ПО приведена на рис. 1. Приложение имеет многопоточную архитектуру. Захват пакетов и управление сетевым интерфейсом осуществляется библиотекой netmap, которая управляется потоком receiver. Receiver подготавливает пакеты и передает данные из циклического буфера netmap в промежуточный буфер приложения, организованный в оперативной памяти (RAM). Запись данных из промежуточного буфера на дисковую подсистему осуществляется потоком writer. Для достижения максимальной производительности учитывается «географическая» привязка PCI-express-слотов, каждому сетевому интерфейсу средствами ОС назначается отдельное ядро центрального процессора.

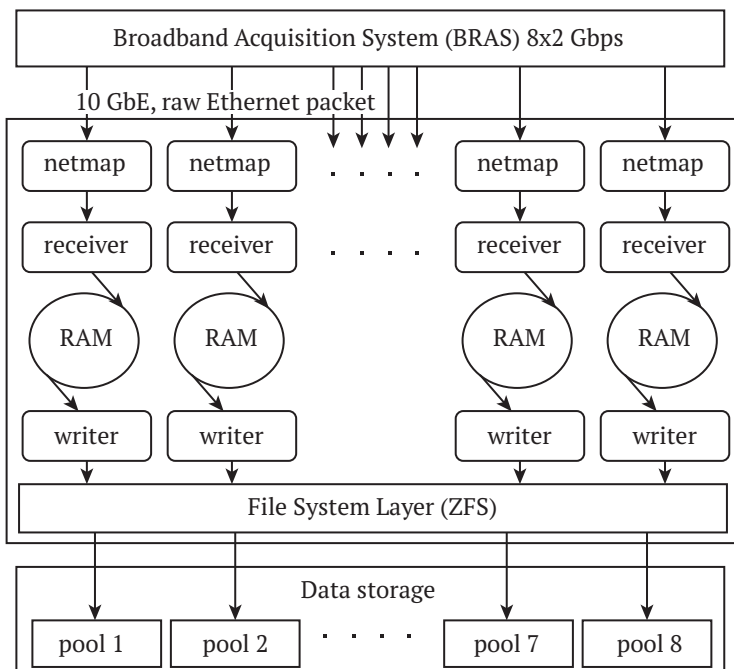


Рис. 1. Структурная схема работы программного обеспечения

## Результат работы ПО pkt-carpes по записи регулярных часовых РСДБ-сессий

В табл. 1 представлены результаты работы ПО в регулярных наблюдениях на радиотелескопах РТ-13 РСДБ-комплекса «Квазар-КВО» за период 2016–2019 гг.

Таблица 1

Результаты записи данных наблюдений ПО pkt-carpes за период 2016–2019 гг.

Год	Код станции	Число сессий	Число сессий с потерей кадров	Число сессий с потерей кадров на первом скане	Объем, ТБ
2016	Bv	1225	56	6	1414
	Zv	1223	116	17	1418
2017	Bv	2046	11	6	2620
	Zv	2032	207	28	2612
2018	Bv	2182	45	6	3230
	Zv	2172	242	188	3176
2019	Bv	631	9	1	850
	Zv	631	67	51	853
	Sw	86	46	12	132

Из таблицы видно, что только в небольшом количестве записанных сканов появляются незначительные потери данных, что не влияет на результаты обработки на корреляторе.

В 2019 г. был проведен ряд экспериментов по захвату и записи потока данных 4 каналов ШСПС на скорости 8 Гбит/с от каждого канала. Каждый канал генерировал поток данных в течение 30 мин по схеме: 30 сек — генерация, 30 сек — пауза. В результате ряда оптимизаций удалось добиться записи потока данных без потерь кадров на протяжении всего сеанса.

## Заключение

Разработанное в ИПА РАН ПО rkt-caprec высокоскоростного потока РСДБ-данных на сетевых интерфейсах в среде ОС FreeBSD с файловой системой ZFS обеспечивает регистрацию потока 8 каналов ШСПС на скорости 16 Гбит/с без потерь и ошибок, что подтверждено результатами многолетних РСДБ-наблюдений на радиотелескопах РТ-13 РСДБ-комплекса «Квазар-КВО».

Осуществлено успешное тестирование режима записи потока данных 4 каналов от ШСПС на скорости 32 Гбит/с, которое возможно использовать для перспективных задач.

В дальнейшем планируется доработать ПО для актуальных версий ОС FreeBSD.

Работа выполнена с использованием оборудования ЦКП/УНУ «Радиоинтерферометрический комплекс «Квазар-КВО».

## Литература

1. Rizzo L. The netmap project [Electronic resource]. — URL: <https://github.com/luigirizzo/netmap> (accessed: 21.10.2019).
2. Ипатов А. В., Гаязов И. С., Смоленцев С. Г. и др. РСДБ-система нового поколения // Труды ИПА РАН. — СПб.: ИПА РАН, 2013. — Вып. 27 — С. 216–222.
3. Маршалов Д. А., Носов Е. В., Гренков С. А., Бердников А. С., Федотов Л. В. и др. Технический облик многофункциональной системы преобразования сигналов для радиотелескопов // Труды ИПА РАН. — СПб.: ИПА РАН, 2017. — Вып. 43. — С. 95–103.
4. Cappallo R., Ruszczyk C., Whitney A. Mark 6: Design and Status. — URL: [http://www.haystack.mit.edu/tech/vlbi/mark6/mark6\\_memo/05-Mark6\\_Design\\_and\\_Status.pdf](http://www.haystack.mit.edu/tech/vlbi/mark6/mark6_memo/05-Mark6_Design_and_Status.pdf) (accessed: 17.10.2019).
5. Salminen T., Mujunen A. Nexpres WP8 FlexBuff [Electronic resource]. — URL: [http://www.jive.nl/nexpres/doku.php?id=nexpres:nexpres\\_wp8](http://www.jive.nl/nexpres/doku.php?id=nexpres:nexpres_wp8) (accessed: 17.10.2019).
6. Verkouter H. VBS\_FS [Electronic resource]. — URL: <http://www.jive.eu/~verkouter/flexbuff/README.vbs> (accessed: 17.10.2019).

# Software for Simultaneous Capture and Recording of a Multi-gigabit VLBI Data Stream as Transmitted over Ethernet

I. A. Bezrukov, A. I. Salnikov, V. A. Yakovlev, A. V. Vylegzhanin

The paper presents rationale for the choice of the structure of data transmission and recording system (DTRS) and describes the software developed for this system based on processing high-speed data flow with netmap. Hardware used to build DTRS (such as server and disk array models, CPUs and NICs, etc.) is described in detail. The results are given of DTRS software performance during the 2016-2019 observations with the RT-13 radio telescopes at Badary, Zelenchukskaya and Svetloe observatories of the Quasar VLBI network.

**Keywords:** recording and storing data, data transfer, e-VLBI.