Аннотация

В работе представлена структура локальной вычислительной сети Пущинского научного центра и Пущинской Радиоастрономической Обсерватории АКЦ ФИАН в том виде, как она выглядит на текущий момент.

Рассмотрены различные уровни архитектуры опорной сети, ее конфигурация, а также представлено оборудование и технологии, используемые для реализации тех или иных задач сети ПНЦ. Представлена подробная информация по интеграции базовой сети в региональную и глобальную вычислительные сети.

Abstract

This paper describes the physical topology of the campus area network of the Pushchino science center of the Russian academy of science, including the LAN of the Pushchino radioastronomical observatory.

We also discuss different aspects of the backbone's implementation, its configuration, as well as configuration and functions of the networks' core servers. This paper also contains detailed information about integration of the core network into the rest of the Internet.

For reference the text contains product names of the communication hardware used.

Локальная вычислительная сеть Пущинского научного центра.

Думский Д.В.[(1)], Исаев Е.А.[(1,2)], Самодуров В. А.[(1,2)].

(1) - Пущинская радиоастрономическая обсерватория астрокосмического центра ФИАН

(2) - Национальный исследовательский университет "Высшая школа экономики"

Структура сети ПНЦ РАН

Вычислительная сеть Пущинского научного центра (Рис.1) логически разделена на три уровня и состоит из ядра, коммутаторов распространения и коммутаторов уровня доступа. Ядро сети и коммутатор распространения расположены в институте математических проблем биологии (ИМПБ) и представляют собой сервер выполняющий роль маршрутизатора и высокопроизводительный коммутатор Allied Telesis AT-X900-24XS с дополнительным модулем AT-XEM12T, главная цель которого быстрый транспорт. Маршрутизатор *octopus.psn.ru* в свою очередь обеспечивает применение политик безопасности, агрегацию и маршрутизацию в VLAN, определяет широковещательные домены, управляет протоколом динамической маршрутизации и выполняет некоторые другие функции. В качестве операционной на сервере работает система Debian GNU/Linux версии 6.0, состоящая из свободного программного обеспечения с открытым исходным кодом.

Коммутаторы уровня доступа, назначение которых подключение конечных устройств, защита от колец в сети (STP) и широковещательных штормов, подключены к Allied Telesis оптоволоконными и медными (в зависимости от расстояния между оборудованием) линиями связи и распределены по институтам и некоторым другим объектам имеющим отношение к инфраструктуре ПНЦ. В частности Пущинская Радиоастрономическая обсерватория и институты ИМПБ, ИТЭБ, ИБК, ИБФМ, ИФХиБПП, ИБП, Институт Белка, а также ППНЦ и городская больница подключены такими коммутаторами к центру сети в ИМПБ (Рис.1).

Подключение локальной сети ПНЦ к Глобальной сети Интернет на канальном уровне осуществляется через расположенный в ФИБХ Дата Центр, принадлежащий компании Stack Data Network (stack.ru), которая предоставляет в

аренду канал шириной 100 Мбит/сек (Рис. 2), подключенный к Президиуму РАН на московском телефонном узле связи №9 (М9) с использованием технологии многопротокольной коммутации по меткам (MPLS).

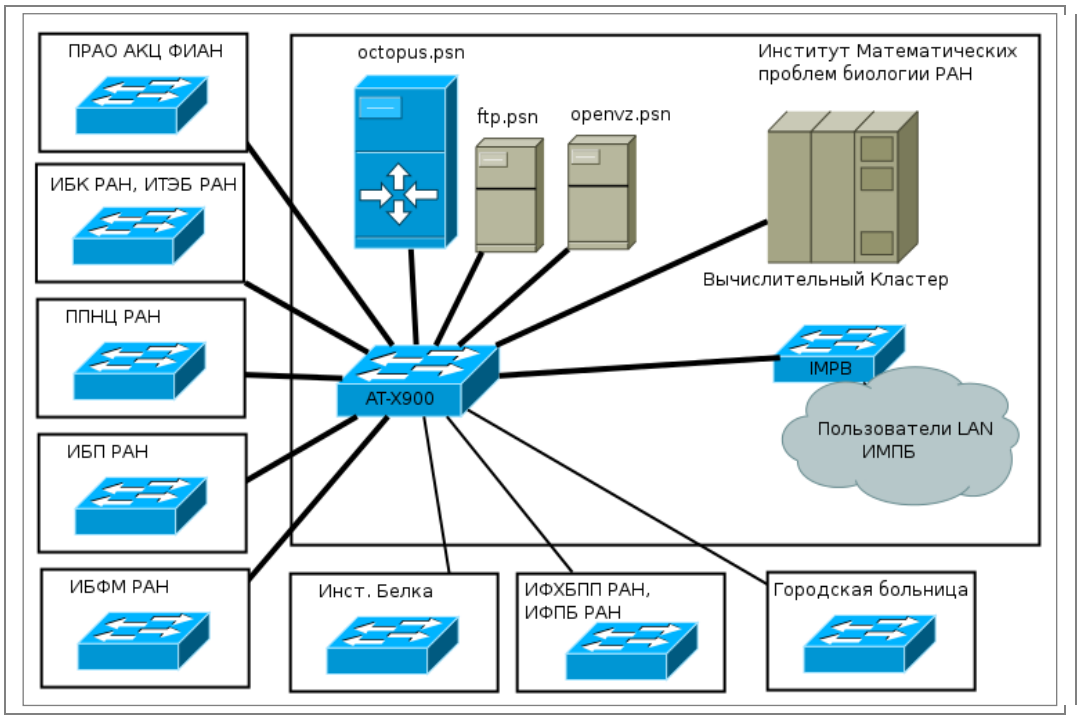

Рисунок 1: Схема локальной вычислительной сети ПНЦ

Канал связи с Президиумом проведен по маршруту ФИБХ-ППНЦ-ИМПБ, и из центра сети в ИМПБ трафик маршрутизируется по всей остальной части вычислительной сети. В дополнение к каналу Президиума РАН в Stack Data Network арендован канал шириной в 1 Гбит/сек выделенный для связи ПРАО с АКЦ ФИАН в Москве. Канал предназначен для передачи научной и телеметрической информации в рамках международного проекта космического радиотелескопа «Радиоастрон» и проведен по маршруту АКЦ-М9-ФИБХ-ПРАО.

Организация сети ПНЦ на третьем уровне показана на схеме Рис. 3. Сеть ПНЦ обладает собственной автономной системой (АС) с номером AS9056. АС ПНЦ является многоинтерфейсной т.е. имеет соединения с двумя интернет провайдерами, что позволяет данной АС оставаться подключенной к Интернету даже в случае обрыва соединения с одним из интернет-провайдеров. При этом транзитный трафик от одного интернет провайдера к другому в АС такого типа запрещен. В качестве основного провайдера используется подключение к авто-

номной системе Президиума РАН (AS3058), в качестве резервного — подключение к городскому провайдеру Интернет ООО «Информационные технологии и электронные коммуникации» (ИТЭК). Сети организаций ПНЦ подключены к центральному маршрутизатору либо напрямую, используя его в качестве шлюза, либо используют собственные маршрутизаторы и подключены способом «точка-

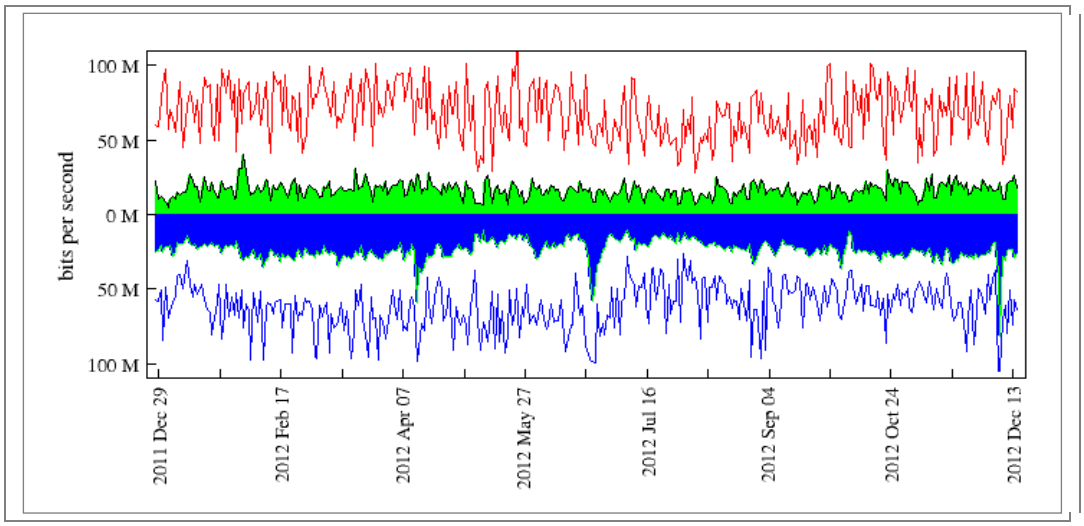

Рисунок 2: Статистика использования канала Интернет за год. На графиках сверху входящий; снизу - исходящий трафик. Кривые с заполнением - усредненные значения, без заполнения - пиковые значения загрузки канала.

точка» (P2P) позволяющим объединять в ip сеть не более двух устройств.

За АС ПНЦ в базе RIPE NCC закреплен блок ip-адресов 194.149.64.0/21, который в свою очередь разбит на 8 подсетей /24, которые распределены по институтам и в зависимости от их потребностей либо используются полностью, либо разделены на более мелкие подсети (Рис. 3). Отдельно выделена сеть внеполосного управления коммутаторами. В целях безопасности доступ к ней разрешен только с центрального маршрутизатора

Вычислительная сеть включает в себя, помимо коммутаторов и маршрутизатора, сервера виртуализации, высокопроизводительный вычислительный кластер и сервера хранения данных.

Программное обеспечение сети ПНЦ РАН

Для анонсирования автономной системы и блока адресов ПНЦ автономным системам Президиума РАН и второго провайдера, а также для получения

маршрутов в Интернет по протоколу BGP (англ. Border Gateway Protocol, протокол граничного шлюза — основной протокол динамической маршрутизации в Интернете) на центральном маршрутизаторе *octopus.psn.ru* используется программное обеспечение (ПО) BIRD (англ. BIRD Internet Routing Daemon http://bird.network.cz/). BIRD поддерживает работу протоколов BGPv4, RIPv2, OSPFv2, OSPFv3 и виртуального протокола Pipe для обмена маршрутами между различными таблицами маршрутизации. Для всех протоколов реализована работа с IPv6.

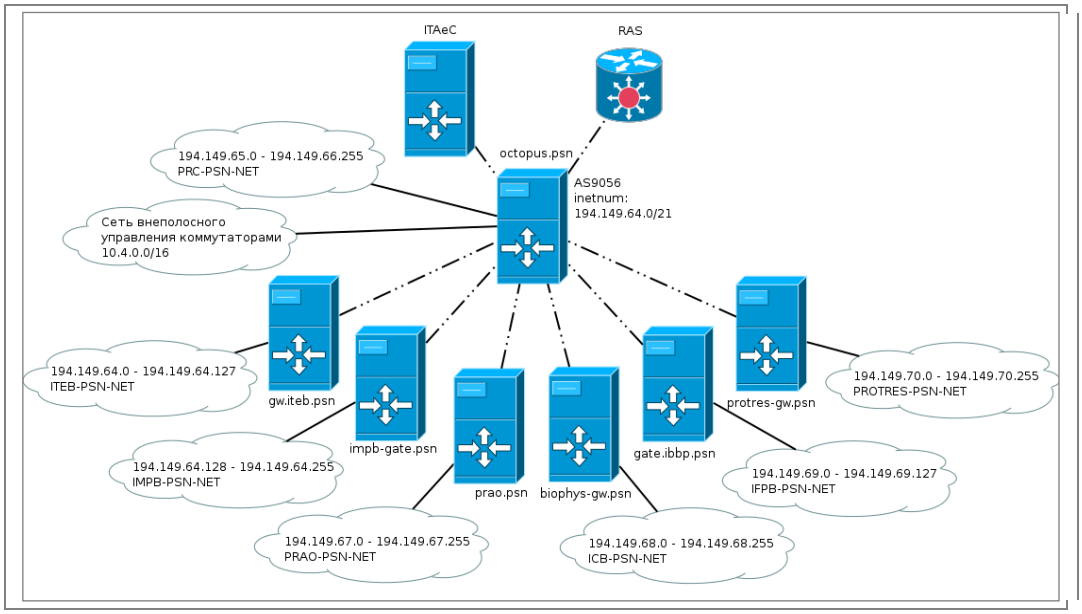

Рисунок 3: Схема третьего уровня ЛВС ПНЦ

Сервер *octopus.psn.ru* служит также пунктом сбора и визуализации статистической информации по техническому состоянию сети, загруженности и доступности каналов связи. Сюда же приходят логи с коммутаторов уровня доступа, и происходит их опрос по протоколу snmp для получения разнообразной диагностической информации о состоянии сети на втором уровне. Статистика работы коммутаторов и загруженность каналов отображается в виде графиков на специально выделенном для этих целей веб-сайте. Системные уведомления с этого и других серверов доставляются администраторам сети почтовым сервером (MTA) exim4.

На серверах виртуализации используется технология OpenVZ, работающая на уровне операционной системы, которая базируется на ядре Linux. OpenVZ позволяет на одном физическом сервере запускать множество изолированных

копий одной операционной системы, называемых «виртуальными частными серверами» (VPS) или «виртуальными средами» [1] (Рис. 4).

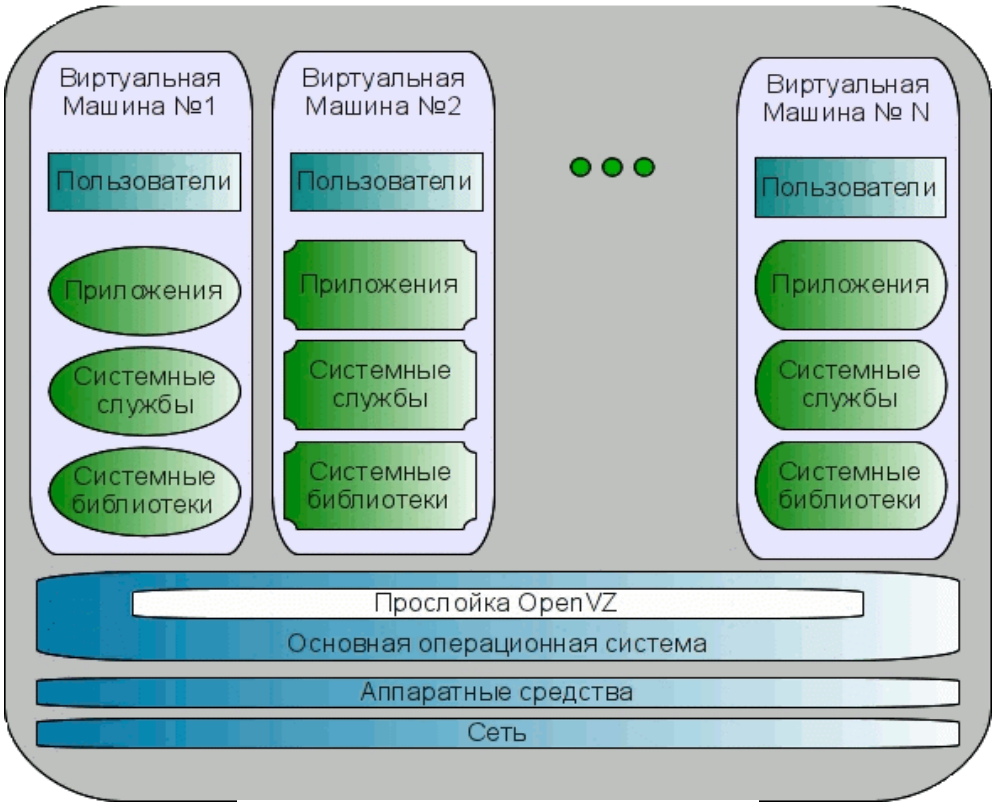

Рисунок 4: Диаграмма виртуализации

В качестве таких «гостевых систем» на сервере OpenVZ в настоящее время работают 25 VPS среди них почтовый сервер (*mail.psn.ru)*, сервера баз данных (*mysql-server, postgresql-server),* сервер DNS, веб сервера ПНЦ (*www.psn.ru*) и некоторых институтов, а также сайт совета молодых ученых и специалистов ПНЦ РАН (*smuis.psn.ru*). Каждому виртуальному серверу выделены соответствующие его потребностям ресурсы основного физического сервера и ведется мониторинг их использования. Такой подход позволяет достичь значительной экономии вычислительных ресурсов по сравнению с использованием нескольких физически раздельных серверов, в то же время, предоставляя лучшую безопасность и надежность работы по сравнению с эксплуатацией всех этих служб под одной операционной системой.

На сетевом уровне VPS связаны между собой отдельной сетью (Servers DMZ LAN) не имеющей выхода в интернет (Рис. 5). В этой же сети одним из своих интерфейсов присутствует сервер для обновлений операционной системы, представляющий собой репозиторий ОС Linux Debian (ftp.psn.ru), а также сервер синхронизации времени (ntp.psn.ru).

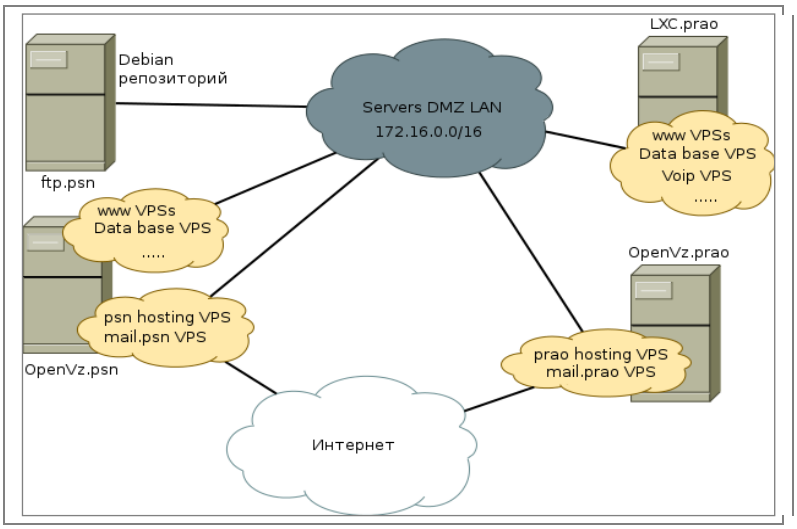

Рисунок 5: Сеть серверов виртуалилизации

Вычислительный кластер, расположенный в ИМПБ состоит из 11 двухпроцессорных узлов на базе микропроцессоров Intel Xeon 5650 и 5640, смонтированных в общую стойку. Общее количество вычислительных ядер 124шт., 248 гигабайт оперативной памяти. В качестве внутренней сетевой среды между узлами используется Infiniband. Программное обеспечение строится на базе свободно распространяемой OS GNU/Linux Debian и интерфейса передачи сообщений MPI как основного средства реализации параллельных вычислений. В качестве основного программного средства организации параллельных вычислений используется OpenMPI версии 1.4.3. Также доступна альтернативная реализация MPI - MPICH2 версии 1.4. Производительность вычислительного кластера на текущий момент составляет порядка 900 Гфлоп.

Структура ЛВС ПРАО АКЦ ФИАН

Отдельно рассмотри устройство локальной вычислительной сети Пущинской радиоастрономической обсерватории Астрокосмического центра ФИАН им.Лебедева (ПРАО АКЦ ФИАН).

Основная особенность организации сети ПРАО АКЦ ФИАН состоит в территориальной распределенности организации. Здесь на территории общей площадью порядка 1.5 км^2 располагаются четыре радиотелескопа, лабораторный корпус и ряд других объектов, включенных в общую вычислительную сеть. Основное количество компьютеров сосредоточено в лабораторном корпусе и небольшое число разбросано по нескольким удаленным друг от друга на 100-1000

м. корпусам (корпуса РТ-22, корпус ДКР-1000, БСА, служба времени, буферный дата-центр, хозблоки) (Рис. 6).

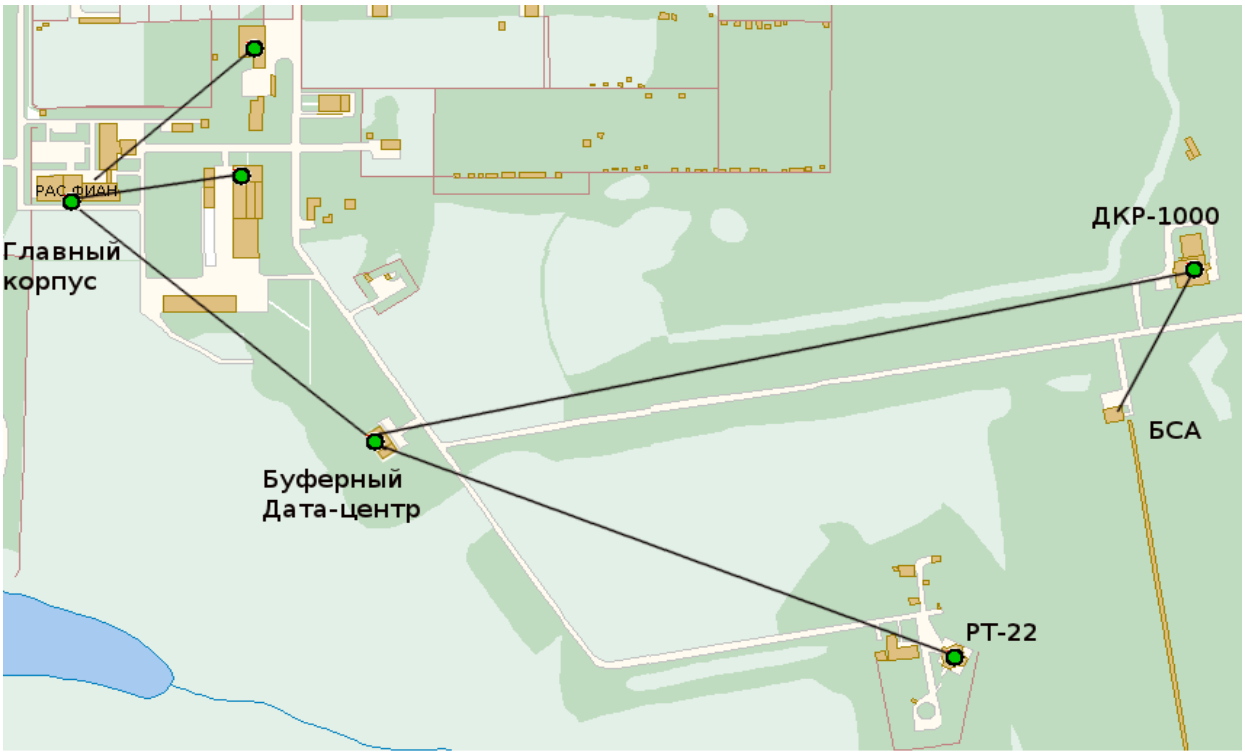

Рисунок 6: Территориальное распределение сети ПРАО АКЦ ФИАН

Сеть ПРАО также можно логически разбить на три уровня. Ядро сети программный маршрутизатор на базе сервера с операционной системой GNU Linux Debian. На уровне распространения используется управляемый коммутатор D-Link DGS-3420-26SC, оснащенный портами 1 Гигабит SFP и 10 Гигабит SFP+, обладающий высокой пропускной способностью. К нему подключены одномодовыми оптическими линиями коммутаторы уровня доступа, распределенные по территории обсерватории.

Отказоустойчивость сети особенно актуальна в свете функционирования в составе ПРАО автоматизированных наблюдательных комплексов [3], поэтому ядро сети и коммутатор уровня распределения размещены в буферном дата-центре оборудованном системой кондиционирования и изолированном от внешней среды в отдельном модуле с фальш-полом. В помещении буферного центра размещены стойки, в которые установлены сервера и источники бесперебойного питания. В задачи буферного дата-центра помимо центра сети обсерватории входит поддержание канала передачи данных и телеметрической информации в проекте космического радиотелескопа “Радиоастрон” и резервное хранение данных. Канал связывает между собой станцию слежения РТ-22, буферный дата-

центр, расположенные на территории обсерватории и центр обработки научной информации (ЦОНИ) АКЦ ФИАН в Москве. Канал пропускной способностью в 1 Гб/с как уже упомяналось ранее предоставлен компанией Stack Data Network и проброшен с использованием технологии MPLS (Multiprotocol Label Switching) до М9, где подключен через провайдера ИКИ к АКЦ ФИАН (Рис. 7). Перед вводом канала в эксплуатацию были проведены исследования его пропускной способности, а также подбиралось программное обеспечение и протоколы для передачи данных на скоростях близких к максимально возможным для используемых серверов приема и передачи данных. Канал полностью изолирован от глобальной сети Интернет и оборудование осуществляющее передачу находится в сети с приватными адресами. В настоящее время полученные с космического радиотелескопа экспериментальные научные данные передаются по каналу на сервера хранения данных в ЦОНИ, часть этих данных одновременно поступает в резервное хранилище буферного дата-центра ПРАО емкостью 20 Тб.

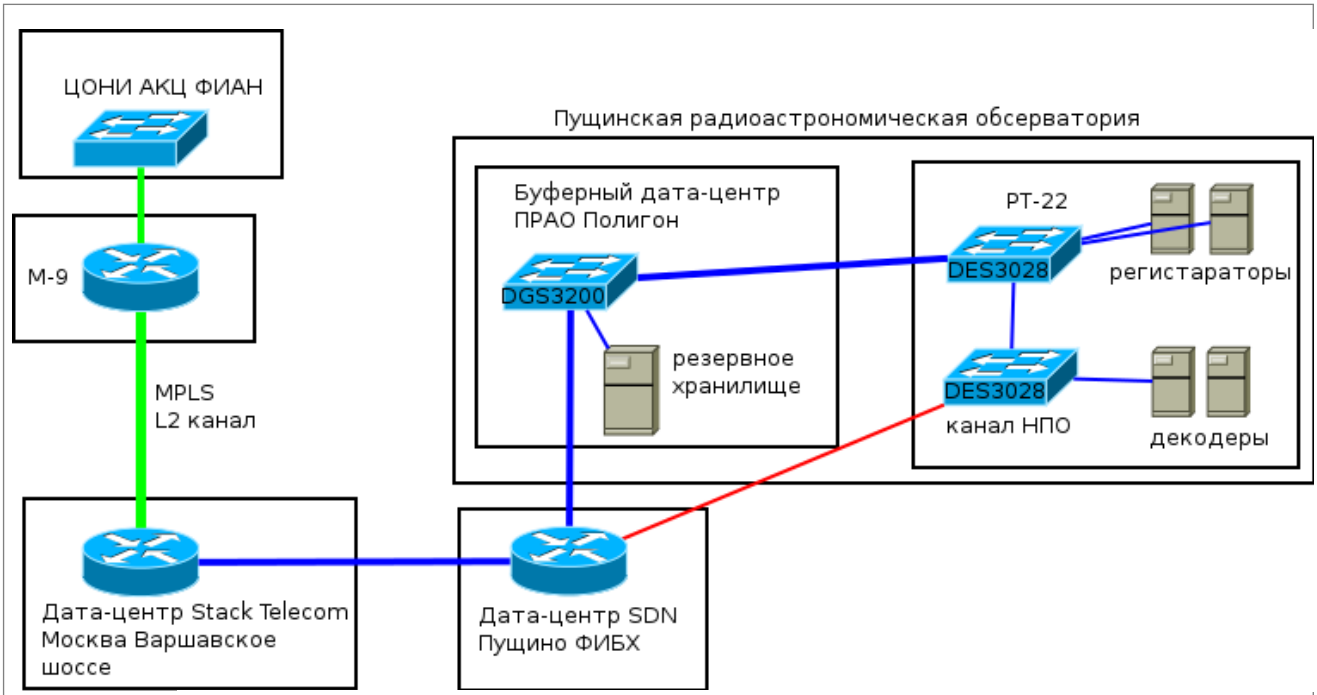

Рисунок 7: Схема подключения каналов связи проекта "Радиоастрон"

Выход ЛВС ПРАО в Интернет обеспечивается с помощью оптоволоконного соединения между ИМПБ РАН и центральным маршрутизатором ПРАО (1 Гбит/с по SFP).

Администрирование и мониторинг серверов осуществляется в штатном режиме через удаленный вход в систему по протоколу SSH v.2. Использование первой версии этого протокола не допускается в связи с обнаруженными в ней

уязвимостями [5]. На случай аварийных ситуаций буферный дата-центр оборудован консолью (монитор и клавиатура), к которой через KVM-переключатель подсоединены все сервера.

В качестве основной операционной системы для серверов используется свободно распространяемая операционная система с открытым исходным кодом Debian GNU/Linux 6.0, известная своей надежностью, стабильностью работы, и сертифицированная по системе CGL (Carrier-Grade Linux).

Для организации удобной рабочей среды ЛВС ПРАО предоставляет ряд важных сетевых сервисов, среди них:

- DHCP-сервер и DNS-сервер;
- WWW-сервера;
- сервер баз данных;
- сервер электронной почты;
- сервер точного времени NTP.

Большинство сервисов, необходимых для эффективной работы сети ПРАО распределены по нескольким серверам, так как их эксплуатация лишь на одном сервере может отрицательно сказываться на уровне его загрузки, а также неизбежно ведет к снижению уровня безопасности. На сервере в ядре сети помимо статической маршрутизации задействованы службы динамической конфигурации узла (DHCP); преобразования сетевых адресов (NAT) и межсетевой экран. Остальные сервисы вынесены на два отдельных сервера виртуализации и функционируют в виде изолированных VPS серверов. К таким службам относятся служба времени (ntp.psn.ru), разрешения доменных имен (ns.prao.ru), сервера обрабатывающие http-запросы от клиентов (веб-сервера), электронная почта и база данных. На серверах виртуализации наряду с платформой OpenVz используется и относительно новая система нативных контейнеров lxc. Данная система не требует дополнителных манипуляци с ядром операционной системы и больше отвечает требованиям безопасности т.к. дистрибутивное ядро обновляется гораздо чаще, чем ядро OpenVz.

На отдельном сервере реализовано сетевое хранилище информации объемом в 31.5 Тб управляемое Open-E Data Storage Software V6 и предоставляющее пользователям доступ к своему дисковому массиву по протоколам http, ftp, sftp и smb.

В ЛВС ПРАО службы, доступные из Интернета (к примеру, веб-сервера, DNS и электронная почта) логически отделены от внутренней сети локальных пользователей, и выделены в так называемую демилитаризованную зону. Вхо-

дящих трафик из Интернета может достичь и повлиять только на службы находящиеся в ДМЗ, и не пропускается во внутреннюю сеть.

Назначением DHCP-сервера является автоматизация процесса выдачи абонентским устройствам сети IP адресов из выделенного пула приватных адресов по заранее заданным администратором ЛВС правилам. Это позволяет существенно сэкономить время, необходимое для подключения к сети новых персональных компьютеров, а также снизить вероятность возникновения конфликтов IP-адресов, в случае ошибок возникающих при назначении адреса вручную.

Маршрутизатор обеспечивает доступ пользователям локальной сети ПРАО к ресурсам Интернет и сети ПНЦ РАН, а также защиту находящихся в ЛВС ПРАО машин от несанкционированного доступа и сетевых атак с помощью сетевого экрана. Маршрутизация и управление трафиком осуществляется с помощью пакета программ iproute2 (*http://lartc.org/howto*); политики сетевого экрана реализованы с помощью модуля ядра Linux iptables []. (*http://www.netfilter.org/projects/iptables*).

В качестве MTA (mail transfer agent: программа, взаимодействующая с другими почтовыми серверами в сети Интернет по протоколу SMTP для обмена электронной почтой и ее маршрутизации) в ЛВС ПРАО используется созданный в Кембриджском университете почтовый сервер Exim4; в качестве MDA (mail delivery agent - программа доставки почты) используется программа высокого уровня защищенности Dovecot, что позволяет пользователям получать доступ к своей электронной почте по протоколам POP3 или IMAP. В качестве альтернативного способа работы с электронной почтовой системой предусмотрен веб-интерфейс roundcube. Для фильтрации спама и вирусов используются программы, greylist и clamav. Для защиты от подбора паролей пользователей fail2ban.

В роли программного обеспечения для веб-серверов выступает Lighttpd: быстрый и защищённый, а также соответствующий стандартам веб-сервер, в котором благодаря асинхронной обработке сетевых соединений, в отличие от другого широко распростарненного веб-сервера Apache, загруженность сервера при доступе к файлам на диске не зависит от количества текущих соединений. Веб-сервера распределны по VPS на двух серверах виртуализации. Каждая VPS обслуживает отдельный сайт среди них www.prao.ru — основной сайт обсерватории, сайты научных конференций и школ ПРАО, и астрономических баз данных и результатов наблюдений. В качестве сервера баз данных выбрана свободная объктно-реляционная система управления PostgreeSQL.

Для трансляции IP-адресов в доменные имена в обсерватории использует-

ся DNS-сервер, построенный на основе программного пакета djbdns, включающего в себя набор утилит для обслуживания и разрешения DNS зон.

Сервер времени используется для автоматической синхронизации времени серверов и рабочих станций по протоколу NTP (Network Time Protocol, RFC 1305 [6]). Данный протокол синхронизирует время в географически распределенных сетях по порту 123 протокола UDP (User Datagram Protocol). Протокол NTP поддерживает множественные избыточные источники данных о времени, что обеспечивает его постоянную синхронизацию, и позволяет установить время на локальной машине с точностью до одной миллисекунды.

Последние два года на обсерватории успешно внедрена и используется ip-телефония. На ее основе организована связь с удаленными от основного здания корпусами обслуживающими радиотелескопы ДКР-1000, БСА и РТ-22, а также мастерскими (Рис. 8). Необходимость передачи голосового трафика посредством ethernet возникла в следствии выхода из строя аналоговых телефонных линий до удаленных корпусов. При принятии решения о создании альтернативного способа телефонизации сыграли свою роль уже существующие оптические линии передачи.

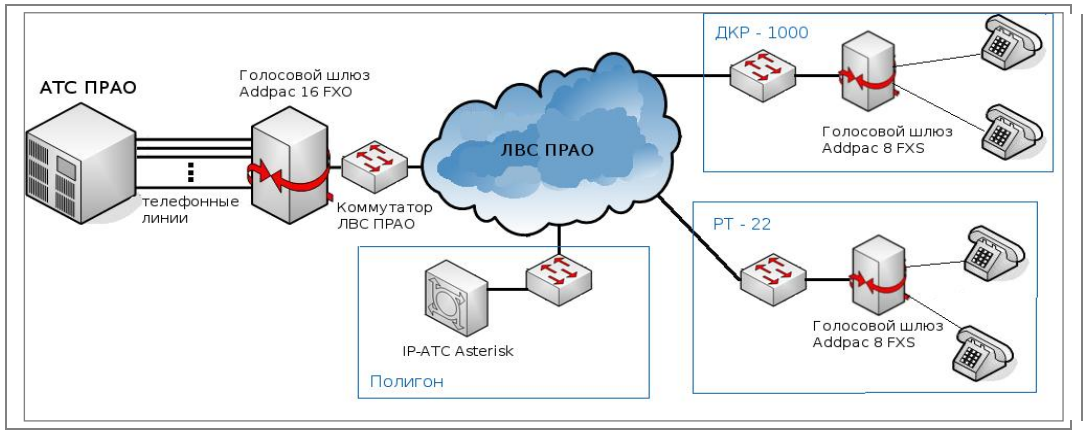

Рисунок 8: Интеграция ip-телефонии в локальную сеть ПРАО

В главном корпусе ПРАО функционирует система цифровой коммутации (СЦК) «ЭЛКОМ», обеспечивающая выход на городские телефонные линии. Для передачи голосового трафика внутри сети аналоговые телефонные линии подключены в голосовые шлюзы AddPac ADD-AP2120-16O и Nateks VC-220 осу-

ществляющие преобразование аналогового сигнал в цифровой и передачу его посредством ethernet в локальную сеть ПРАО (Рис. 5). Для маршрутизации полученного голосового трафика используется свободная программная АТС Asterisk с открытым исходным кодом, разработанная компанией Digium. Asterisk в комплексе с необходимым оборудованием обладает всеми возможностями классической АТС, поддерживает множество VoIP-протоколов и предоставляет богатые функции управления звонками. На стороне конечных пользователей установлены голосовые шлюзы осуществляющие обратное преобразование голосового ip-трафика в аналоговый сигнал на обычные телефонные аппараты. Голосовые шлюзы по протоколу SIP (*Session Initiation Protocol* - протокол установления сеанса)[7] регистрируются на сервере Asterisk конфигурация которого позволяет направлять поступившие на СЦК «ЭЛКОМ» звонки на телефоны конечных пользователей, и позволяет этим пользователям звонить на внешние номера используя соответствующие линии.

Интеграция ЛВС ПНЦ РАН и ЛВС ПРАО АКЦ ФИАН в глобальную вычислительную сеть

Данный раздел посвящен описанию BGP-связей ПНЦ и смежных сетей, а также общей структуре сегмента Интернета, к которому они подключены. Информация представлена в виде таблиц и автоматически построенных графов (таблица 1, рис. 9).

Пиринг	Импорт	Экспорт
AS2643 IHEP-SU AS	AS2643	AS9056 AS13161
AS2683 RADIO-MSU	ANY	AS9056 AS2643
AS3058 RAS-AS Russian Academy of Sciences	ANY ANY AND NOT FLTR BOGONS	AS9056 AS9056
AS21453 FLEX-AS	ANY	AS9056
AS41783 ITAEC-AS	AS41783 AS16083 ANY AND NOT FLTR-BOGONS AS41783 AS16083	AS9056 AS9056

Таблица 1: Пиринговые связи системы AS9056

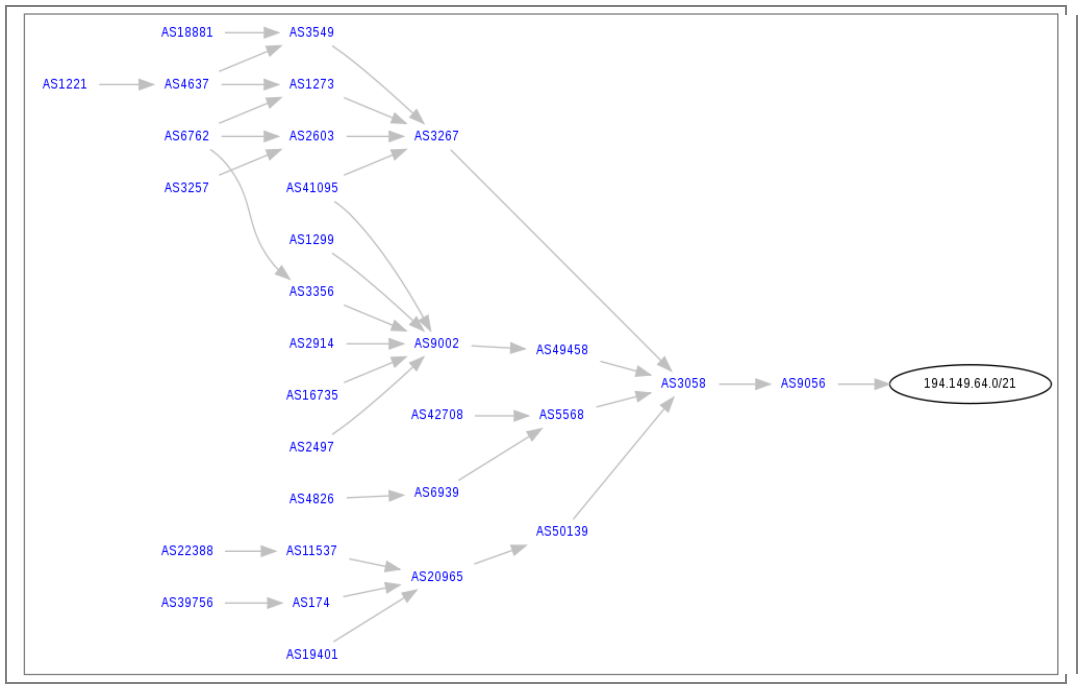

Рисунок 9: Визуализация графа маршрутизации внешнего трафика AS9056 - автономной системы ПНЦ РАН (подсеть 194.149.64.0/21).

Список литературы

[1] K Kolyshkin. Virtualization in Linux. *http://mirrors.unbornmedia.com/openvz/doc/openvz-intro.pdf*, 2006.

[2] IEEE Std. 802.1Q-2005, Virtual Bridged Local Area Networks, ISBN 0-7381-3662-X.

[3] В.В. Китаев. Распределенная система обработки и сбора данных ПРАО АКЦ ФИАН. II Базовая локальная вычислительная сеть; препринт ФИАН №52, Москва, 1997.

[4] IEEE 802.3 LAN/MAN CSMA/CD (Ethernet) Access Method, *http://standards.ieee.org/getieee802/802.3.html*

[5] CORE SDI S.A. SSH Insertion attack.1998. US CERT Vulnerability Note VU#13877

[6] David L. Mills, University of Delaware. Network Time Protocol (Version 3). Specification, Implementation and Analysis. IETF RFC1305. March 1992.

[7] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks,

M. Handley, E. Schooler. SIP: Session Initiation Protocol. IETF RFC 3261. June 2002.

Список аббревиатур

ПРАО: Пущинская радиоастрономическая обсерватория

АКЦ: астрокосмический центр

ФИАН: физический институт академии наук

ПНЦ: Пущинский научный центр ИТЭБ: институт теоретической и экспериментальной биофизики

ИБК: институт биофизики клетки

ИФХиБПП: институт физико-химическеских и биологических проблем почвоведения

ИФПБ: институт фундаментальных проблем биологии

ИБП: институт биологического приборостроения

ИМПБ: институт математических проблем биологии

ФИБХ: филиал института биоорганической химии

ИБ: институт белка

ИБФМ: институт биохимии и физиологии микроорганизмов

ППНЦ: президиум Пущинского научного центра

ЛВС: локальная вычислительная сеть