

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

Федеральное государственное автономное образовательное  
учреждение высшего образования  
ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

Институт математики, механики и компьютерных наук  
имени И. И. Воровича

Направление подготовки  
01.03.02 — Прикладная математика  
и информатика

РЕАЛИЗАЦИЯ РАСПРЕДЕЛЁННОГО ОТКАЗОУСТОЙЧИВОГО  
ПРОСТРАНСТВА КОРТЕЖЕЙ СРЕДСТВАМИ ЯЗЫКА RUTRON

Выпускная квалификационная работа  
на степень бакалавра

Студентки 4 курса  
В. М. Ложкиной

Научный руководитель:  
старший преподаватель кафедры информатики и вычислительного  
эксперимента В. Н. Брагилевский

Допущено к защите:  
заведующий кафедры ИВЭ \_\_\_\_\_ В. С. Пилиди

Ростов-на-Дону  
2018

# Содержание

Введение . . . . .	3
1. Распределённые вычислительные системы и отказоустойчи- вость . . . . .	5
2. BTS: Византийское пространство кортежей . . . . .	7
2.1. Системная модель . . . . .	7
2.2. Модуль <code>secondary_functions</code> . . . . .	11
2.3. Модуль <code>client</code> . . . . .	12
2.4. Модуль <code>BTS_infrastructure</code> . . . . .	13
2.5. Модуль <code>BTS_server</code> . . . . .	17
2.6. Операция записи <code>out</code> . . . . .	19
2.7. Операция чтения <code>rd</code> . . . . .	19
2.8. Операция чтения <code>inp</code> . . . . .	21
Заключение . . . . .	27
Список литературы . . . . .	27
Приложение . . . . .	29

# Введение

В современном мире распределённые системы являются основой различных объектов, например, веб-сервисов и пиринговых сетей. Поскольку обмен сообщениями в системе осуществляется по ненадёжным каналам связи, сообщения, посылаемые системе или самой системой, могут просматриваться, перехватываться и подменяться. Это приводит к неправильному функционированию, отказу отдельных компонент или системы в целом. Поэтому в распределённых системах поддержке безопасности уделяется особое внимание.

Для повышения надёжности системы случайные и умышленные сбои интерпретируются как византийские ошибки (в терминах задачи о византийских генералах). В этом случае можно использовать соответствующие отказоустойчивые техники, которые смогут защитить систему и от случайных сбоев, и от умышленных вторжений.

В данной работе рассматривается византийское пространство кортежей — открытая распределённая устойчивая к византийским ошибкам система, основанная на пространстве кортежей без распределённой памяти, в которой процессы взаимодействуют путём обмена сообщениями.

Раздел 1 включает в себя определение основных терминов, использованных в данной работе, таких как распределённая система, отказоустойчивость и задача византийских генералов. Раздел 2 посвящён реализации византийского пространства кортежей. Он включает в себя восемь подразделов. В подразделе 2.1 описана системная модель византийского пространства кортежей (BTS). В подразделах 2.2-2.5 рассматривается программная реализация распределённой отказоустойчивой системы BTS на языке программирования Python, каждый модуль реализации описывается отдельно. Особое внимание уделяется операциям манипулирования данными, описанным в подразделах 2.6-2.8, поскольку именно их реализация обеспе-

чивает отказоустойчивость системы. Результаты работы реализованной системы, а также их анализ представлены в разделе ??.

# **1. Распределённые вычислительные системы и отказоустойчивость**

Распределённая вычислительная система — это набор независимых компьютеров, реализующий параллельную обработку данных на многих вычислительных узлах [1]. С точки зрения пользователя этот набор является единым механизмом, предоставляющим полный доступ к ресурсам. Существует возможность добавления новых ресурсов и перераспределения их по системе, возможность добавления свойств и методов, но информация об этих событиях скрыта от пользователя.

Одной из важнейших характеристик распределённых систем является отказоустойчивость. Отказоустойчивость — это свойство системы сохранять работоспособность в том случае, если какие-либо составляющие её компоненты перестали правильно функционировать [1]. Компоненты системы могут стать неработоспособны по различным причинам, например, из-за технологических сбоев или атак безопасности.

Большинство современных распределённых систем имеют характеристики открытых систем. Открытая распределённая система предполагает использование служб, вызов которых требует стандартного синтаксиса и семантики [2]. Такая система может иметь неизвестное количество ненадёжных и неоднородных участников, к тому же участникам не нужно быть активными одновременно (свойство разъединённости во времени) и не обязательно что-то знать друг о друге (свойство разъединённости в пространстве). Связь между узлами распределённой системы является ненадёжной (может прерываться, что повлечёт за собой потерю сообщений), обмен сообщениями может происходить не мгновенно, а с существенной задержкой. Кроме того, любой узел системы может отказать или быть выключен.

чен в любой момент времени. Все эти факторы неизбежно приводят к неправильному функционированию системы.

Один из способов улучшить её надёжность — это интерпретировать случайные или умышленные неполадки как византийские ошибки (в терминах задачи византийских генералов), тогда использование отказоустойчивых техник сможет сделать координационную составляющую системы отказоустойчивой и для случайных сбоев, и для умышленных вторжений.

Задача византийских генералов — это задача взаимодействия нескольких удалённых абонентов, получивших сообщения из одного центра, причём часть этих абонентов, в том числе центр, могут быть предателями, то есть могут посылать заведомо ложные сообщения с целью дезинформирования. Нахождение решения задачи заключается в выработке единой стратегии действий, которая будет являться выигрышной для всех абонентов [3].

Формулировка задачи состоит в следующем. Византийская армия представляет собой объединение некоторого числа легионов, каждым из которых командует свой генерал, генералы подчиняются главнокомандующему армии Византии. Поскольку империя находится в упадке, любой из генералов и даже главнокомандующий могут быть заинтересованы в поражении армии, то есть являться предателями. Генералов, не заинтересованных в поражении армии, будем называть верными. В ночь перед сражением каждый из генералов получает от главнокомандующего приказ о действиях во время сражения: атаковать или отступить. Таким образом, имеем три возможных исхода сражения:

- благоприятный исход: все генералы атакуют противника, что приведёт к его уничтожению и победе Византии.
- промежуточный исход: все генералы отступят, тогда противник не будет побеждён, но Византия сохранит свою армию.

- неблагоприятный исход: некоторые генералы атакуют противника, некоторые отступят, тогда Византийская армия потерпит поражение.

Так как главнокомандующий тоже может оказаться предателем, генералам не следует доверять его приказам. Однако, если каждый генерал будет действовать самостоятельно, независимо от других генералов, то вероятность наступления благоприятного исхода становится низкой. Таким образом, генералам следует обмениваться информацией между собой для того, чтобы прийти к единому решению.

В данной работе рассматривается распределённая система под названием «Византийское пространство кортежей» (BTS: Byzantine Tuple Space) [4]. Все сбои, происходящие в этой системе, интерпретируются как византийские ошибки, для устранения которых используются определённые отказоустойчивые техники.

## 2. BTS: Византийское пространство кортежей

### 2.1. Системная модель

Системная модель византийского пространства кортежей BTS предполагает бесконечное число процессов-клиентов  $\Pi = \{p_1, p_2, \dots\}$ , которые коммуницируют со множеством из  $n$  серверов  $U = \{s_1, s_2, \dots, s_n\}$  при помощи обмена сообщениями.

Будем полагать, что случайное число клиентов и связка серверов из  $f \leq \left\lfloor \frac{n-1}{3} \right\rfloor$  штук могут быть подвержены византийским ошибкам: они могут произвольным образом отклоняться от их спецификаций и работать в сговоре, чтобы изменить поведение системы. Такие процессы будем называть неисправными, а правильно работающие процессы — корректными.

Основа распределённой системы BTS — это пространство кортежей, которое также является ядром языка программирова-

ния Linda [5], предназначенного для построения эффективных параллельных программ. Кorteж — это структура данных, представляющая собой неизменяемый список фиксированной длины, элементы которого могут относиться к различным типам данных. Два corteжа  $t_1$  и  $t_2$  считаются идентичными, если совпадают их длины, а также типы и значения соответствующих полей. Хранилище corteжей, в котором доступ к элементам может осуществляться параллельно, называется пространством corteжей [6].

Каждый сервер  $s \in U$  при запуске получает на вход имя файла, в котором хранится содержимое пространства corteжей. Это необходимо для создания сервером  $s$  локальной копии пространства corteжей  $T_s$ . Кроме того, создаётся изначально пустое множество corteжей для удаления  $R_s$ . Будем полагать, что идентичных corteжей в пространстве не существует, тогда к двум описанным множествам corteжей могут быть применены стандартные операции над множествами.

Для искусственной имитации неисправных серверов будем при запуске подавать им на вход файл с неправильным пространством corteжей. Таким образом, множества  $T_{s_1}$  и  $T_{s_2}$ , принадлежащие корректному серверу  $s_1$  и неисправному серверу  $s_2$  соответственно, будут иметь пустое пересечение, что повлечёт за собой конфликты и позволит проверить отказоустойчивость системы.

Каждый сервер  $s$  реализует три операции манипулирования данными (corteжами):

- *out* — запись corteжа в пространство corteжей.
- *rd* — недеструктивное чтение corteжа.
- *in* — деструктивное чтение (извлечение) corteжа.

Заметим, что операции *rd* и *in* не являются блокирующими в отличие от их аналогов в языке программирования Linda.



Определим такое понятие, как шаблон кортежа — это кортеж, некоторые поля которого неопределены и не представляют важности. Будем говорить, что кортеж соответствует шаблону, если длина кортежа равна длине шаблона и определённые в шаблоне поля совпадают по типу и значению с соответствующими полями кортежа.

Операция записи *out* принимает в качестве входного параметра кортеж, все поля которого определены. Операции чтения *rd* и *in* принимают в качестве входного параметра шаблон кортежа, по которому производится поиск соответствующих ему кортежей в пространстве.

Благодаря наличию операций чтения/записи, пространство кортежей можно рассматривать как разновидность распределённой памяти: например, одна группа процессов записывает данные в пространство, а другая группа извлекает их и использует в своей дальнейшей работе.

Распределённая система не может полагаться на какой-то один конкретный узел, так как в случае его отказа восстановить систему будет невозможно [7]. Это означает, что выполнение любой операции манипулирования кортежами не может производиться только на одном из имеющихся узлов. Для решения этой проблемы разобьём множество  $U$  на кворумы и получим кворум-систему, тогда каждая операция чтения/записи будет выполняться в соответствующем кворуме, что обеспечит правильное функционирование системы в случае отказа  $f$  узлов.

Кворум-система представляет собой множество кворумов серверов  $Q \in 2^U$ , в котором каждая пара кворумов из  $Q$  пересекается на достаточно многих серверах и всегда есть кворум со всеми корректными серверами. Существование пересечений между кворумами позволяет совершенствовать протоколы чтения/записи, позволяя поддерживать целостность разделённой переменной, даже если эти операции были выполнены в разных кворумах системы.

Будем полагать, что  $n > 3f + 1$ . Разделим кворумы на два типа ( $Q = Q_r \cup Q_w$ ):

- кворумы чтения ( $Q_r \in \mathcal{Q}_r$ ), мощность каждого кворума  $|Q_r| = \left\lceil \frac{n + f + 1}{2} \right\rceil$  серверов,
- кворумы записи ( $Q_w \in \mathcal{Q}_w$ ), мощность каждого кворума  $|Q_w| = \left\lceil \frac{n + f + 1}{2} \right\rceil + f$  серверов.

Так как  $|Q_r| + |Q_w| > n$ , то можно ожидать, что полученное при чтении значение будет наиболее актуальным, поскольку хотя бы один из узлов кворума  $|Q_r|$  также участвует в выполнении операции записи.

Взаимодействие клиентов с системой происходит посредством вспомогательной промежуточной координационной инфраструктуры, как показано на рисунке 1.

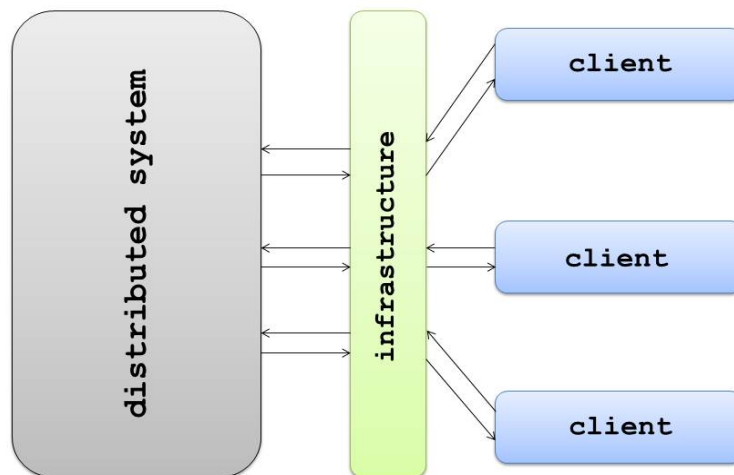


Рисунок 1 — Взаимодействие клиентов с распределённой системой

Когда клиент  $p$  посылает запрос системе, этот запрос сначала обрабатывается координационной инфраструктурой и только после этого отправляется системе. При получении ответа от системы координационная инфраструктура также обрабатывает полученную информацию, а затем отправляет ответ клиенту  $p$ . Поэтому клиентская

часть операций чтения/записи выполняется на стороне координационной структуры, а не на стороне клиента.

Реализация системы BTS состоит из 4 модулей:

- модуль `secondary_functions` включает в себя вспомогательные функции,
- модуль `client` реализует клиентский интерфейс,
- модуль `BTS_infrastructure` реализует деятельность координационной инфраструктуры,
- модуль `BTS_server` реализует деятельность сервера  $s \in U$ .

Рассмотрим реализацию компонент более подробно.

## 2.2. Модуль `secondary_functions`

Модуль `secondary_functions` включает в себя функции сетевого взаимодействия (`get_message`, `send_message`, `connect_to_server`), а также функцию создания и записи пространства кортежей в файл (`create_ts_file`). Реализации данных функций представлены в листинге .1 в приложении.

Функции сетевого взаимодействия используют в своей работе возможности модуля `socket` из стандартной библиотеки Python `??`. Функция `connect_to_server` принимает в качестве входного параметра порт, к которому необходимо подключиться. При удачном подключении возвращается сокет, который можно использовать для обмена сообщениями в дальнейшей работе, иначе возвращается объект `None`, сигнализирующий о неудаче.

Функция `send_message` используется для отправки данных по сети, она имеет два входных параметра: сокет, с помощью которого осуществляется отправка сообщения, и само сообщение, которое требуется отправить. Чтобы преобразовать сообщение к виду, пригодному для записи в сокет, используется функция `dumps` из моду-

ля `pickle` ??, отвечающая за сериализацию объектов. По окончании операции записи в сокет функция `send_message` завершает свою работу.

Функция `get_message` используется для получения данных из сети, она имеет единственный входной параметр — сокет, из которого осуществляется чтение информации. Если полученное сообщение оказалось пустым, то возвращается объект `None`, иначе сообщение десериализуется с помощью функции `loads` из модуля `pickle` и возвращается функцией `get_message`.

Функция `create_ts_file` создаёт пространство кортежей и записывает его в файл. Данная функция используется для того, чтобы в дальнейшем каждый сервер системы при запуске смог прочитать информацию из созданного при помощи `create_ts_file` файла и сконструировать локальную реплику пространства кортежей. Имя файла, объём создаваемого пространства, а также длина и значения полей входящих в него кортежей регулируются входными параметрами функции. Для преобразования созданного пространства кортежей в удобный формат и дальнейшей записи его в файл используется функция `dump` из модуля `json` ??.

### 2.3. Модуль `client`

Интерфейс пользователя для взаимодействия с BTS реализован с помощью класса `Client` из модуля `client` (полный программный код представлен в приложении). Для того, чтобы начать работу, следует создать объект класса `Client`, передав конструктору в качестве входного параметра значение порта, который координационная инфраструктура использует для обмена сообщениями с клиентами.

Далее для исполнения операций чтения/записи нужно вызвать соответствующую функцию: `out_op(t)`, `rd_op( $\bar{t}$ )` или `in_op( $\bar{t}$ )`. Отметим, что операция записи `out_op` в качестве входного параметра принимает кортеж, а операции чтения `rd_op` и `in_op` — шаблон

кортежа. Каждая из этих операций формирует запрос в виде словаря {'op': код операции, 'tup': кортеж} или {'op': код операции, 'temp': шаблон кортежа}, соответствующего её семантике. Сформированный запрос отправляется координационной инфраструктуре. Для завершения операции записи ответ системы не требуется, а операции чтения ожидают, когда инфраструктура вернёт запрошенное значение.

Для корректного завершения работы системы в классе `Client` реализована функция `stop_op()`, которая отправляет инфраструктуре запрос о прекращении работы и ожидает ответа об успешном результате операции. По смыслу данная функция не должна являться частью клиентского интерфейса, она была включена в класс `Client` для удобства.

Ввиду возможности возникновения ошибок для контроля исполняемых операций используется модуль `logging` ?? из стандартной библиотеки языка Python, отвечающий за логирование. Создание и настройка конфигурации лога происходит в конструкторе класса `Client` при помощи функции `basicConfig`, устанавливающей имя файла для записи лога, формат выводимых сообщений и уровень логирования. В реализации для вывода информационных сообщений использован уровень `INFO`.

## 2.4. Модуль `BTS_infrastructure`

Координационная инфраструктура системы BTS реализована в модуле `BTS_infrastructure` с помощью класса `BTS_infrastructure` (полный программный код представлен в приложении). Она представляет собой многопоточное приложение, которое обрабатывает запросы клиента и взаимодействует с серверами системы BTS.

Для того, чтобы настроить работу координационной системы, необходимо создать объект класса `BTS_infrastructure` и передать в его конструктор следующие параметры (см. листинг ??):

- порт, который будет прослушивать инфраструктура и на который будут поступать клиентские запросы,
- список портов, каждый из которых будет прослушивать определённый сервер системы BTS,
- количество неисправных серверов (предателей), которые будут порождать конфликты в системе,
- размер кворума записи,
- размер кворума чтения,
- имя файла, в котором хранится множество кортежей для корректных серверов,
- имя файла, в котором хранится множество кортежей для неисправных серверов.

В теле конструктора значения входных параметров присваиваются соответствующим свойствам класса, устанавливается значение флага `THREAD_POOL_ON` (данный флаг используется при обработке клиентских запросов), значение свойства `AMOUNT_OF_CLIENTS` устанавливается в нуль (данное свойство является счётчиком поступивших клиентских запросов), множество серверов `U` разбивается на кворумы записи `QW` и чтения `QR`, а также настраивается конфигурация лога.

После создания объекта класса для запуска работы координационной инфраструктуры необходимо вызвать метод `run` (см. листинг ??). Если изначально количество неисправных серверов было указано неверно, то есть не меньше количества всех имеющихся серверов, то возвращается значение `False`. Иначе работа функции продолжается: происходит запуск корректных и неисправных серверов с помощью метода `start_servers` (см. листинг ??).

Метод `start_servers` имеет три входных параметра:

- два индекса, определяющие диапазон значений в списке портов `self.U`, которые будут прослушивать запускаемые серверы,
- имя файла, которое будет подано выбранным серверам на вход.

Программная реализация сервера системы BTS представлена в модуле `BTS_server`, который будет рассмотрен позже в разделе 2.5, данный скрипт `BTS_server.py` запускается с определёнными параметрами в новом процессе при помощи конструктора класса `Popen` из модуля `subprocess` ?? стандартной библиотеки Python. В скрипт передаются следующие аргументы:

- идентификатор сервера,
- порт, который будет прослушивать сервер,
- имя файла с хранящимся в нём пространством кортежей,
- список портов кворума, в котором состоит сервер.

Информация об исполненных операциях записывается в лог инфраструктуры.

Вернёмся к методу `run`. После запуска серверов создаётся и настраивается сокет, с помощью которого далее будет происходить коммуникация с клиентскими процессами. Будем обрабатывать клиентские запросы при помощи пула потоков `ThreadPoolExecutor` из модуля `concurrent.futures` ?? стандартной библиотеки Python, принимающего в качестве аргумента количество потоков, используемых в работе.

Пока флаг `THREAD_POOL_ON` равен `True`, выполняется следующая последовательность инструкций: ожидается подключение клиента к порту, создаётся соответствующий клиентский сокет, количество полученных клиентских запросов увеличивается на единицу. При удачном завершении описанных инструкций задача с помощью метода `submit` добавляется в очередь для исполнения пу-

лом потоков. Метод `submit` принимает в качестве входных параметров имя функции (`worker`), которую необходимо исполнить в одном из потоков пула, и значения аргументов (клиентский сокет и `AMOUNT_OF_CLIENTS`), с которыми эта функция будет запущена.

Флаг `THREAD_POOL_ON` может изменить своё значение только при получении запроса о прекращении работы от клиента. Если это произошло, то новые задачи перестают добавляться в очередь пула потоков, ожидается завершение всех исполняющихся или ожидающих в очереди задач. После завершения работы пула потоков вызывается метод `stop_servers` (см. листинг ??), при исполнении которого всем запущенным серверам посылается запрос о прекращении работы. На этом работа метода `run` завершается, возвращается значение `True`, свидетельствующее об успешном завершении работы координационной инфраструктуры.

Заметим, что метод `run` является блокирующим, то есть после запуска инфраструктуры и вплоть до её завершения продолжить дальнейшую работу будет невозможно.

Рассмотрим метод `worker` (см. листинг ??) более подробно. Он выполняет функцию получения и обработки запроса клиента, принимает в качестве входных параметром клиентский сокет и идентификатор клиентского процесса. Вспомним, что в функцию `submit` (добавление задачи в очередь пула потоков) в качестве второго аргумента функции `worker` передаётся значение свойства `AMOUNT_OF_CLIENTS`. Данное свойство используется для подсчёта клиентских запросов, будем полагать, что каждый клиентский запрос был послан уникальным клиентом (согласно системной модели BTS множество клиентских процессов бесконечно), тогда в качестве идентификатора клиентского процесса можно использовать номер запроса, что и осуществляется при помощи свойства `AMOUNT_OF_CLIENTS`.

Работу метода `worker` можно разделить на два этапа: получение сообщения от клиента и обработка полученного запроса. Для того, чтобы получить сообщение с запросом от клиента, исполь-



зуется функция `get_message` из описанного в разделе 2.2 модуля `secondary_functions`. После получения запроса определяется код операции, которую необходимо исполнить, вызывается соответствующая функция и при необходимости отправляется ответ клиенту.

При получении запросов на исполнение операций *out*, *rd* или *in* вызываются методы `out`, `rdp` или `inp`, которые будут рассмотрены позже в разделах ??, ?? и ?? соответственно.

При получении запроса на исполнение операции *stop* флаг `THREAD_POOL_ON` устанавливается в значение `False`, после чего прекращается обработка клиентских запросов, система готовится к завершению работы.

## 2.5. Модуль `BTS_server`

Серверная часть системы `BTS` реализована в модуле `BTS_server` (полный программный код представлен в приложении), он представляет собой скрипт, запуск которого приводит в исполнение сервер системы. Как уже было отмечено в разделе 2.4, скрипт запускается со следующими аргументами:

- идентификатор сервера,
- порт, который будет прослушивать сервер,
- имя файла с хранящимся в нём пространством кортежей,
- список портов кворума, в котором состоит данный сервер.

Для обработки аргументов используется библиотека `argparse` ?. Она предоставляет возможности анализа аргументов командной строки, конвертирования строковых аргументов в другие объекты программы и вывода информационных подсказок.

Для мониторинга работы сервера настраивается логирование.

При запуске сервер считывает из полученного файла кортежи и создаёт локальную копию пространства кортежей при помощи функ-

ции `read_from_ts_file` (см. листинг ??). На сервере кортежи хранятся в контейнере `set`, что позволяет исключить наличие идентичных кортежей и сделать проверку на вхождение быстрой, поскольку `set` в качестве базовой структуры данных использует хэш-таблицу ??.

Далее, как и в случае с координационной инфраструктурой, создаётся сокет, через который будет осуществляться коммуникация с инфраструктурой, запускается пул потоков ?? и начинается обработка входящих запросов.

Обработка входящих запросов осуществляется при помощи функции `worker`, программный код которой приведён в приложении в листинге ?. В теле данной функции сначала происходит считывание сообщения, полученного от инфраструктуры, далее по содержанию полученного сообщения определяется тип команды и вызывается соответствующая одноимённая функция. Полный перечень возможных команд приведён ниже:

- `out` - команда операции записи *out*,
- `rd` - команда операции недеструктивного чтения *rd*,
- `in` - команда операции деструктивного чтения *in*,
- `enter` - команда входа в критическую секцию,
- `exit` - команда выхода из критической секции,
- `assert1` - команда первого этапа утверждения удаляемого кортежа,
- `assert2` - команда второго этапа утверждения удаляемого кортежа,
- `stop` - команда о прекращении работы сервера.

Первые три команды связаны с операциями манипулирования кортежами, следующие четыре команды связаны с реализацией операции *in* и будут рассмотрены позже в разделе ?. Последняя команда `stop`

устанавливает флаг `THREAD_POOL_ON` в значение `False`, что приводит к прекращению работы пула потоков, а затем — к завершению работы сервера.

## 2.6. Операция записи `out`

Операция  $out(t)$  добавляет кортеж  $t$  в пространство кортежей. На стороне инфраструктуры эта операция реализуется с помощью функции `out(self, t)`, программный код которой приведён в листинге ?? в приложении.

Для того, чтобы добавить кортеж  $t$  в пространство, производится подключение к серверам системы, состоящим в кворуме записи, и в случае удачного подключения отправляется запрос `{ 'op': 'out', 'tup': t }` на добавление кортежа.

Заметим, что при вызове данной функции ждать ответа от серверов нет необходимости. Будем считать, что операция завершится в тот момент, когда все корректные серверы из кворума записи получат кортеж  $t$ .

На стороне сервера операция  $out(t)$  реализуется с помощью функции `out(t)` (программный код приведён в листинге ?? в приложении). При получении запроса на добавление кортеж  $t$  добавляется в пространство кортежей только в том случае, если этот кортеж ранее не был из него удалён (то есть его нет во множестве `RS`). Это необходимо для того, чтобы один и тот же кортеж не был удалён дважды.

## 2.7. Операция чтения `rd`

Операция недеструктивного чтения  $rd(\bar{t})$  в качестве входного параметра принимает шаблон кортежа  $\bar{t}$  и возвращает копию соответствующего шаблону кортежа, не извлекая его из пространства. На стороне инфраструктуры эта операция реализуется с помощью функции `rdp(self, temp, ts=None)`, имеющей два парамет-

ра: шаблон кортежа и объект `ts` (по умолчанию равный `None`), значение которого будет объяснено позже. При получении запроса на исполнение операции `rd` функция `rdp` вызывается с единственным первым параметром. Программный код приведён в листинге ?? в приложении.

Суть работы функции заключается в следующем. Каждому серверу из кворума чтения посылается запрос на исполнение операции  $rd(\bar{t})$ . Ответы от серверов кворума приходят в виде множества подходящих под шаблон  $\bar{t}$  кортежей и записываются в словарь `ts`, имеющий структуру: `port: (socket, set)`, где `port` - порт, который прослушивает конкретный сервер, `socket` - сокет, с помощью которого происходит коммуникация, `set` - множество подходящих кортежей, хранящихся на данном сервере. Заметим, что словарь `ts` может передаваться в функцию уже в готовом виде, этот факт используется в реализации операции  $in(\bar{t})$ , которая будет рассмотрена в разделе ??.

Из всех полученных кортежей выбирается один общий кортеж  $t$ , который располагается как минимум на  $f + 1$  сервере. Чтобы найти кортеж  $t$ , необходимо подобрать такую комбинацию из  $f + 1$  сервера, пересечение множеств подходящих кортежей которых будет непусто. Для этого рассматриваются всевозможные сочетания из  $f + 1$  сервера при помощи функции `combinations` из модуля `itertools` ?. Для нахождения пересечений множеств подходящих кортежей используются методы класса `set` ?. После нахождения кортежа  $t$  он возвращается в качестве выходного параметра. Если подходящего кортежа не нашлось, то возвращается объект `None`, свидетельствующий о том, что операция чтения не увенчалась успехом.

Рассмотрим реализацию функции неdestructивного чтения `rdp(sock, temp)` на стороне сервера, которая принимает два входных параметра: сокет и шаблон кортежа (программный код представлен в листинге ?? в приложении). Для того, чтобы найти подходящие под шаблон кортежи, производится один проход по локальной копии

пространства кортежей  $TS$ , каждый кортеж которого проверяется на пригодность функцией `match`. Если функция `match` вернула значение `True`, то кортеж добавляется во множество подходящих кортежей `temp_set`.

Функция `match(temp, t)` сравнивает длины шаблона и кортежа, в случае совпадения длин сравниваются определённые в шаблоне поля с соответствующими полями проверяемого кортежа. Если длины кортежей оказались разными или какое-то определённое поле шаблона не совпало с соответствующим полем кортежа, то возвращается значение `False`, говорящее о том, что кортеж не подходит под шаблон. Иначе возвращается значение `True`.

Если переданный в функцию `rdp` сокет оказался не равен значению `None`, то полученное множество `temp_set` отправляется инфраструктуре в качестве ответа. Иначе оно возвращается в качестве выходного параметра. Такая архитектура необходима для реализации операции  $in(\bar{t})$ , которая будет рассмотрена в следующем разделе.

## 2.8. Операция чтения `inr`

Операция деструктивного чтения  $in$  имеет более сложную реализацию, поскольку один кортеж не может быть удалён из пространства двумя различными вызовами  $in$ . Этот факт подразумевает использование критических секций для процессов, пытающихся удалить один и тот же кортеж, что, во-первых, обеспечит невозможность удаления кортежа двумя процессами одновременно, во-вторых, позволит всем процессам получить доступ к ресурсу в некоторое время.

Помимо критических секций необходимо использовать алгоритм достижения консенсуса ??, чтобы в результате исполнения операции  $in$  из пространства был удалён правильный кортеж. Ярким представителем таких алгоритмов является алгоритм Paxos, предложенный Лесли Лампортом ??.

Рахос — это алгоритм для решения задачи консенсуса в сети ненадёжных вычислителей. Компоненты распределённой системы можно разделить на три группы ??:

- Заявитель (Proposer) — выдвигает «предложения» (какие-то значения), которые либо принимаются, либо отвергаются в результате работы алгоритма.
- Акцептор (Acceptor) — принимает или отвергает «предложение» Заявителя, согласует своё решение с остальными Акцепторами, уведомляет о своём решении Узнающих. Если Акцепторами было принято какое-либо значение, предложенное Заявителем, то оно называется утверждённым.
- Ученик (Learner) — запоминает решение Акцепторов, принятое в результате работы алгоритма консенсуса.

Компоненты распределённой системы могут принадлежать сразу нескольким группам, описанным выше, и вести себя и как Заявитель, и как Акцептор, и как Ученик. Такое распределение ролей в системе гарантирует следующее ??:

- Только предложенное Заявителем значение может быть утверждено Акцепторами.
- Акцепторами утверждается только одно значение из всех предложенных Заявителями значений (возможно, противоречивых).
- Ученик не сможет узнать об утверждении какого-либо значения вплоть до того момента, пока оно действительно не будет утверждено.

Распределим роли среди компонент системы BTS следующим образом:

1. Множество Заявителей состоит только из координационной инфраструктуры. Такой набор гарантирует наличие корректного Заявителя.

2. Множество Акцепторов составляют все серверы системы.
3. Множество Учеников включает в себя все серверы системы и инфраструктуру.

Рассмотрим реализацию операции  $in(\bar{t})$  на стороне инфраструктуры, представленную методом `inp(self, pid, temp)` (программный код приведён в листинге ??). Данная функция принимает два параметра: идентификатор клиентского процесса и шаблон кортежа. В первую очередь инфраструктура пытается войти в критическую секцию системы для удаления подходящего под шаблон `temp` кортежа. Функция входа в критическую секцию `enter_r(self, pid, temp)` совмещена с операцией неструктивного чтения  $rd(\bar{t})$  (см. листинг ?? в приложении), это необходимо для того, чтобы уменьшить количество обращений к системе и, как следствие, ускорить исполнение операции  $in$ . По окончании выполнения функции `enter_r` все серверы из кворума чтения уже разрешили войти в критическую секцию, а инфраструктурой уже был выбран конкретный кортеж  $t$  для удаления (с помощью метода `rdp`).

Если не удалось выбрать подходящий кортеж  $t$ , то вызывается метод `exit_r(self, pid, temp)`, которая посылает серверам запрос на выход из критической секции. После этого метод `inp` возвращает значение `None`, свидетельствующее о том, что подходящего для удаления кортежа нет в пространстве.

Если же выбрать кортеж  $t$  всё-таки получилось, то необходимо выяснить, можно ли удалить этот кортеж из пространства. Для этого вызывается метод `propose_learn(self, pid, temp, t, ts)`, в ходе исполнения которого серверам системы отправляется предложение об удалении кортежа  $t$ , после чего от каждого сервера приходит ответ о принятом решении. На основе ответов серверов выбирается самый популярный, его значение возвращается в качестве выходного параметра. Заметим, что в случае наличия  $f$  и менее неисправных серверов в качестве ответа от сервера может прийти либо кортеж  $t$

(значит он уже был удалён из пространства), либо объект `None` (удаления не произошло), поскольку ни один сервер не является Заявителем. Таким образом, метод `propose_learn` симулирует деятельность инфраструктуры как Заявителя и Ученика.

На момент выходы из метода `propose_learn` все серверы уже сделали выбор, какой именно кортеж удалять. Если функция `propose_learn` вернула значение, равное  $t$ , то операция *in* считается успешно завершённой и возвращается кортеж  $t$ . В противном случае необходимо произвести ту же последовательность действий, начиная со входа в критическую секцию. Данный процесс продолжается до того момента, пока серверами не будет одобрено предложенное инфраструктурой значение или пока в пространстве не окажется подходящих для удаления кортежей.

Заметим, что выход из критической секции происходит только в случае неудачного исхода операции удаления, а в остальных случаях этого не происходит. Такая реализация необходима для избежания блокировок, выход из критической секции на каждом сервере осуществляется локально, и при следующей попытке инфраструктуры войти в критическую секцию не возникает проблем.

Теперь рассмотрим реализацию операции *in* на стороне сервера. Исполнение данной операции на стороне инфраструктуры начинается с попытки входа в критическую секцию, рассмотрим, что происходит на сервере в этот момент. При получении запроса на вход в критическую секцию вызывается функция `enter_r(sock, pid, temp)` (см. листинг ?? в приложении), выполняющая две функции: добавление запроса в очередь и конструирование множества кортежей, подходящих под шаблон `temp`.

Когда инфраструктура запрашивает доступ к критической секции, происходит захват ресурса — типа шаблона кортежа. Тип шаблона кортежа — это кортеж, поля которого равны значениям типов соответствующих полей шаблона. Если тип шаблона уже захвачен другим вызовом `enter_r`, то запрос добавляется в очередь, соответству-



ющую данному ресурсу. Иначе происходит захват типа шаблона и исполняются инструкции, отвечающие за конструирование множества кортежей, подходящих под шаблон, с помощью функции `rdp`.

Далее инфраструктура посылает серверам запрос на удаление конкретного кортежа. При получении данного запроса вызывается функция `inp(sock, pid, temp, tup)` с аргументами: сокет, идентификатор клиентского процесса, шаблон кортежа, кортеж, предложенный для удаления. Если клиентский процесс с данным идентификатором `pid` находится в критической секции, начинается исполнение алгоритма Рахос, реализованного в функции `рахос` (см. листинг ?? в приложении).

В реализации используется стандартный Рахос-алгоритм ?. Принятие решения происходит в два этапа. На первом этапе все серверы получают от инфраструктуры кортеж, который необходимо удалить. Если данный кортеж есть в пространстве кортежей, то он считается утверждённым на первом этапе, иначе утверждается значение `None`. Далее все серверы обмениваются утверждёнными на первом этапе значениями, в итоге у каждого сервера имеется вектор утверждённых значений. Это осуществляется при помощи функции `ассепт1(p, temp, server_port, tup)` (см. листинг ?? в приложении), которая помещает полученные от серверов значения в соответствующий контейнер. На рисунке 2 представлена графическая интерпретация алгоритма Рахос, в которой четвёртый сервер является неисправным и посылает сообщения с неверной информацией.

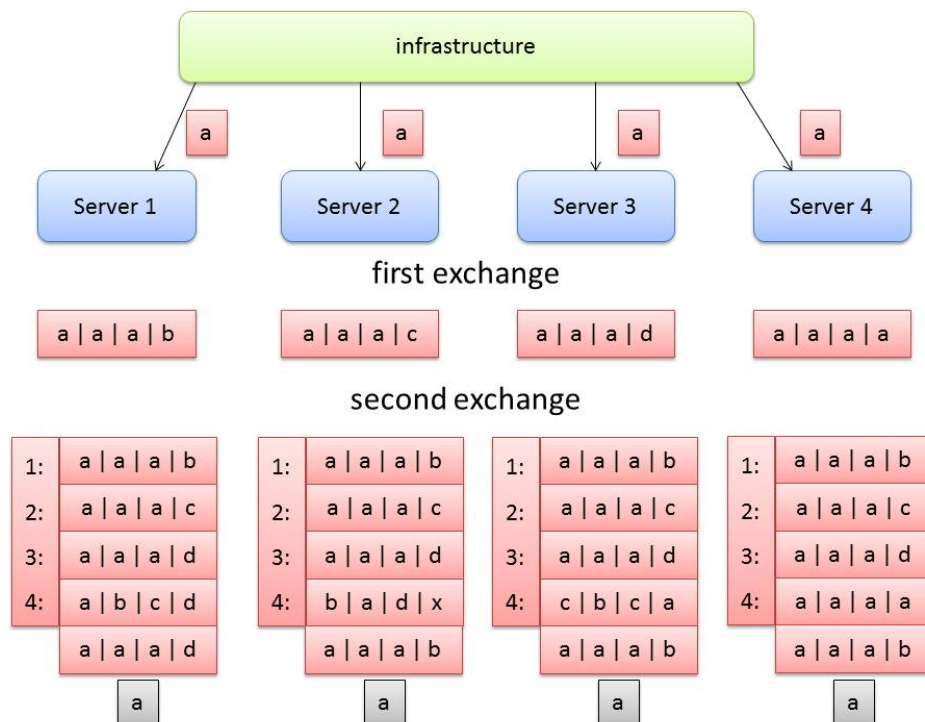


Рисунок 2 — Алгоритм Рахос

Далее следует второй этап: серверы снова обмениваются полученными векторами друг с другом, в итоге у каждого сервера имеется матрица утверждённых значений, как показано на рисунке 2. Данные манипуляции реализованы при помощи функции `assert2(p, temp, server_port, tup_dict)` (см. листинг ?? в приложении). Далее в каждом столбце матрицы выбирается наиболее часто встречающееся значение, оно помещается в результирующий вектор. В результирующем векторе снова находится значение, которое встречается чаще остальных. Это значение является утверждённым на втором этапе. Таким образом Акцепторами принимается решение. Поскольку все сервера являются и Акцепторами, и Учениками, они уже знают, какое решение было принято, единственный Ученик, который не знает о принятом решении, это инфраструктура, поэтому ей отправляется сообщение о том, какой кортеж был удалён из пространства кортежей.

После завершения работы функции `рахос` возвращённое ею значение удаляется из пространства кортежей. Происходит локальное освобождение ресурса при помощи функции `exit_r(temp, rid=None)`, удаляющей соответствующий запрос из очереди. На этом завершается выполнение функции `inr`.

Данный результат иллюстрирует благоприятный и промежуточный исходы задачи византийских генералов. При благоприятном исходе из пространства кортежей удаляется кортеж, предложенный клиентом, то есть приказ главнокомандующего (клиента) был получен и изучен всеми генералами (серверами), после чего они единогласно приняли решение следовать приказу. При промежуточном исходе кортеж не удаляется, возвращается `None`, при этом сохраняется целостность пространства кортежей, поскольку на всех корректных серверах удаления не произошло. Иными словами, генералами (серверами) было принято единогласное решение не доверять приказу главнокомандующего (клиента) и отступить (не производить удаление), сохранив при этом армию (целостность пространства кортежей).

## Заключение

## Список литературы

1. *Таненбаум Э., Стеен М. ван.* Распределённые системы. Принципы и парадигмы. — СПб : Питер, 2003. — ISBN 5-272-00053-6.
2. *Косяков М. С.* Введение в распределенные вычисления. — 2014. — URL: <https://books.ifmo.ru/file/pdf/1551.pdf> (дата обр. 13.02.2018).
3. Wikipedia, the free encyclopedia : Byzantine fault tolerance. — URL: [https://en.wikipedia.org/wiki/Byzantine\\_fault\\_tolerance](https://en.wikipedia.org/wiki/Byzantine_fault_tolerance) (дата обр. 16.04.2018).

4. *Bessani A. N., Silva Fraga J. da, Lung L. C.* BTS: A Byzantine fault-tolerant tuple space // Proceedings of the 2006 ACM symposium on Applied computing. — Dijon, France : ACM. — С. 429—433. — ISBN 1-59593-108-2. — DOI: 10.1145/1141277.1141377.
5. Wikipedia, the free encyclopedia : Linda (coordination language). — URL: <https://ru.wikipedia.org/wiki/Linda> (дата обр. 16.11.2016).
6. Wikipedia, the free encyclopedia : Tuple space. — URL: [https://en.wikipedia.org/wiki/Tuple\\_space](https://en.wikipedia.org/wiki/Tuple_space) (дата обр. 16.04.2018).
7. *Кленнман М.* Высоконагруженные приложения. Программирование, масштабирование, поддержка. — СПб : Питер, 2018. — ISBN 978-5-4461-0512-0.

# Приложение

---

## Листинг .1. Модуль secondary\_functions

---

```
1 from socket import socket, AF_INET, SOCK_STREAM, error
2 import pickle
3 import json
4
5 def get_message(sock):
6     sock.settimeout(50)
7     msg = b""
8     tmp = b""
9     while True:
10         try:
11             tmp = sock.recv(1024)
12         except error:
13             break
14         if len(tmp) < 1024:
15             break
16         msg += tmp
17     msg += tmp
18     if len(msg) > 0:
19         return pickle.loads(msg)
20     else:
21         return None
22
23 def send_message(sock, msg):
24     sock.send(pickle.dumps(msg))
25
26 def connect_to_server(port):
27     sock = socket(AF_INET, SOCK_STREAM)
28     i = 50
29     while i:
30         try:
31             sock.connect(('localhost', port))
32         except ConnectionRefusedError:
33             i -= 1
34         else:
35             return sock
36     return None
37
38 def create_ts_file(file_name, init_num, amount=500, length=100):
39     with open(file_name, 'w') as f:
40         list_of_tuples = list()
41         for i in range(init_num * amount, (init_num + 1) * amount
42             ):
43             list_of_tuples.append(tuple(j for j in range(i, i +
44                 length)))
45         json.dump(list_of_tuples, f)
```

---

---

## Листинг .2. Класс Client, модуль client

---

```
1 import logging
2 from secondary_functions import get_message, send_message,
   connect_to_server
3
4 class Client(object):
5
6     def __init__(self, port):
7         self.port = port
8         logging.basicConfig(filename='client_log.txt', level=
           logging.DEBUG, format="%(asctime)s - %(message)s")
9
10
11     def out_op(self, tup):
12         sock = connect_to_server(self.port)
13
14         if sock is not None:
15             send_message(sock, {'op': 'out', 'tup': tup})
16             sock.close()
17             logging.info('OUT: ' + str(tup))
18         else:
19             logging.info('OUT_OP: cannot connect to server')
20
21
22     def rd_op(self, temp):
23         sock = connect_to_server(self.port)
24
25         if sock is not None:
26             send_message(sock, {'op': 'rdp', 'temp': temp})
27             resp = get_message(sock)
28             sock.close()
29             logging.info('RD: ' + str(resp['resp']))
30
31         if resp is None:
32             return resp
33         else:
34             return resp['resp']
35     else:
36         logging.info('RD_OP: cannot connect to server')
37         return None
```

---

---

**Листинг .3.** Класс `Client`, модуль `client` (продолжение)

---

```
1  def in_op(self, temp):
2
3      sock = connect_to_server(self.port)
4
5      if sock is not None:
6          send_message(sock, {'op': 'inp', 'temp': temp})
7          resp = get_message(sock)
8          sock.close()
9          logging.info('In: ' + str(resp['resp']))
10
11         if resp is None:
12             return resp
13         else:
14             return resp['resp']
15
16     else:
17         logging.info('IN_OP: cannot connect to server')
18
19         return None
20
21  def stop_op(self):
22
23      sock = connect_to_server(self.port)
24
25      if sock is not None:
26          send_message(sock, {'op': 'stop'})
27          resp = get_message(sock)
28          sock.close()
29          logging.info('STOP_OP: ' + str(resp['resp']))
30
31         if resp is None:
32             return resp
33         else:
34             return resp['resp']
35
36     else:
37         logging.info('STOP_OP: cannot connect to server')
38
39         return None
```

---

---

**Листинг .4. Класс BTS\_infrastructure, модуль BTS\_infrastructure**

---

```
1 from secondary_functions import *
2 from subprocess import Popen, CalledProcessError, PIPE
3 from concurrent.futures import ThreadPoolExecutor
4 from itertools import combinations
5 from collections import Counter
6 import logging
7
8 class BTS_infrastructure(object):
9     def __init__(self, port, servers, amount_of_enemies, size_of_QW
10         , size_of_QR, right_file, wrong_file):
11         self.INFRASTRUCTURE_PORT = port
12         self.U = servers
13         self.AMOUNT_OF_SERVERS = len(servers)
14         self.AMOUNT_OF_ENEMIES = amount_of_enemies
15         self.TS_FILE, self.WRONG_TS_FILE = right_file, wrong_file
16         self.THREAD_POOL_ON = True
17         self.AMOUNT_OF_CLIENTS = 0
18         self.QW, self.QR = self.U[0:size_of_QW], self.U[-size_of_QR:]
19         logging.basicConfig(filename='infrastructure_log.txt', level=
20             logging.DEBUG, format="%(asctime)s - %(message)s")
21         logging.info('Infrastructure on port ' + str(self.
22             INFRASTRUCTURE_PORT))
23
24     def start_servers(self, ind1, ind2, file):
25         for i in range(ind1, ind2):
26             if U[i] in self.QR:
27                 quorum = self.QR
28             else:
29                 quorum = [U[i]]
30
31         try:
32             Popen('python BTS_server.py '+str(i)+' '+str(self.U[i])+
33                 '+str(file)+' '+''.join(str(j)+' ' for j in quorum),
34                 stdout=PIPE, stderr=PIPE, shell=True)
35         except CalledProcessError:
36             logging.info("START_SERVERS: cannot start server "+str(self.
37                 U[i]))
38         else:
39             logging.info("START_SERVERS: start server "+str(self.U[i]))
40
41     def stop_servers(self):
42         for i in self.U:
43             s = connect_to_server(i)
44             if s is not None:
45                 send_message(s, {'op': 'stop'})
46                 s.close()
47                 logging.info("STOP_SERVERS: stop server "+str(i))
48             else:
49                 logging.info("STOP_SERVERS: cannot stop server "+str(i))
```

---



---

**Листинг .5.** Класс BTS\_infrastructure, модуль BTS\_infrastructure  
(продолжение)

---

```
1  def worker(self, client, pid):
2      req = get_message(client)
3      try:
4          if req['op'] == 'out':
5              self.out(req['tup'])
6          elif req['op'] == 'rdp':
7              send_message(client, {'resp':self.rdp(req['temp'])})
8          elif req['op'] == 'inp':
9              send_message(client, {'resp':self.inp(pid,req['temp'])})
10         elif req['op'] == 'stop':
11             self.THREAD_POOL_ON = False
12             send_message(client, {'resp': 'ok'})
13         else:
14             logging.info('WORKER: wrong request ' + str(req))
15     except KeyError:
16         logging.info('WORKER: wrong request ' + str(req))
17     client.close()
18
19 def run(self):
20     if self.AMOUNT_OF_ENEMIES >= self.AMOUNT_OF_SERVERS:
21         return False
22
23     self.start_servers(0,self.AMOUNT_OF_ENEMIES,self.
24         WRONG_TS_FILE)
25     self.start_servers(self.AMOUNT_OF_ENEMIES,self.
26         AMOUNT_OF_SERVERS,self.TS_FILE)
27
28     s = socket(AF_INET, SOCK_STREAM)
29     s.bind('', self.INFRASTRUCTURE_PORT))
30     s.listen(1)
31
32     with ThreadPoolExecutor(4) as pool:
33         while self.THREAD_POOL_ON:
34             try:
35                 s.settimeout(30)
36                 client_s, client_addr = s.accept()
37                 self.AMOUNT_OF_CLIENTS += 1
38             except error:
39                 pass
40             else:
41                 pool.submit(self.worker, client_s, self.
42                     AMOUNT_OF_CLIENTS)
43
44     self.stop_servers()
45     s.close()
46     logging.info('RUN: stop working')
47     return True
```

---

---

**Листинг .6.** Класс BTS\_infrastructure, модуль BTS\_infrastructure (продолжение)

---

```
1  def enter_r(self, pid, temp):
2      ts = {i: [None, None] for i in self.U} # port: (socket, set)
3
4      for i in self.U:
5          ts[i][0] = connect_to_server(i)
6          if ts[i][0] is not None:
7              send_message(ts[i][0], {'op': 'enter', 'pid': pid, 'temp'
8                  : temp})
9              logging.info('ENTER_R: send ' + str(temp) + ' to server '
10                  + str(i))
11          else:
12              if i in self.QR:
13                  return tuple([None, None])
14              else:
15                  ts.pop(i)
16
17      qr = set()
18
19      for i in ts.keys():
20          resp = get_message(ts[i][0])
21          ts[i][0].close()
22
23          if resp is None or resp['resp'] != 'go':
24              logging.info('ENTER_R: wrong respond from server ' + str(
25                  i))
26              ts[i][1] = None
27          else:
28              ts[i][1] = resp['ts']
29              qr.add(i)
30
31      if not set(self.QR).issubset(qr):
32          return tuple([None, None])
33
34      return tuple([self.rdp(temp, ts), ts])
35
36  def exit_r(self, pid):
37      for i in self.U:
38          sock = connect_to_server(i)
39
40          if sock is not None:
41              send_message(sock, {'op': 'exit', 'pid': pid})
42              sock.close()
43          else:
44              logging.info('EXIT_R: cannot exit cs with pid: ' + str(
45                  pid))
```

---

---

**Листинг .7.** Класс BTS\_infrastructure, модуль BTS\_infrastructure (продолжение)

---

```
1  def out(self, t):
2  for i in self.QW:
3      sock = connect_to_server(i)
4      if sock is not None:
5          send_message(sock, {'op': 'out', 'tup': t})
6          logging.info('OUT: send ' + str(t) + ' to server ' + str(i)
7              )
8          sock.close()
9      else:
10         logging.info('OUT: cannot connect to server ' + str(i))
11
12 def inp(self, pid, temp):
13 while True:
14     t, ts = self.enter_r(pid, temp)
15     logging.info('INP: enter with pid ' + str(pid) + ', temp ' +
16         str(temp) + ', t: ' + str(t))
17     if t is None:
18         self.exit_r(pid)
19         logging.info('INP: exit from cs with pid ' + str(pid))
20         return None
21
22     d = self.paxos(pid, t, ts)
23     self.exit_r(pid)
24     logging.info('INP: exit from cs with pid ' + str(pid))
25     if d == t:
26         break
27
28     return t
29
30 def paxos(self, pid, t, ts):
31     for i in ts.keys():
32         ts[i][0] = connect_to_server(i)
33         if ts[i][0] is not None:
34             send_message(ts[i][0], {'op': 'in', 'pid': pid, 'tup': t})
35
36     tup_list = []
37     for i in ts.keys():
38         resp = get_message(ts[i][0])
39         ts[i][0].close()
40         if resp is not None:
41             tup_list.append(resp['resp'])
42         else:
43             tup_list.append(None)
44
45     c = Counter(tup_list)
46     return c.most_common(1)[0][0]
```

---

---

**Листинг .8.** Класс BTS\_infrastructure, модуль BTS\_infrastructure (продолжение)

---

```
1  def rdp(self, temp, ts=None):
2      if ts is None:
3          ts = {i: [None, None] for i in self.U} # port: (socket,
4              set)
5          for i in self.U:
6              ts[i][0] = connect_to_server(i)
7              if ts[i][0] is not None:
8                  send_message(ts[i][0], {'op': 'rd', 'temp': temp})
9                  logging.info('RDP:send '+str(temp)+'to server '+str(i))
10             else:
11                 ts.pop(i)
12
13         qr = set()
14         for i in ts.keys():
15             msg = get_message(ts[i][0])
16             ts[i][0].close()
17             if msg is not None:
18                 ts[i][1] = msg['ts']
19                 qr.add(i)
20             else:
21                 ts[i][1] = None
22
23         logging.info('RDP: receive ' + str(ts[i][1]) + ' from
24             server ' + str(i))
25         if not set(self.QR).issubset(qr):
26             return None
27
28         t = None
29         for i in combinations(ts.keys(), self.AMOUNT_OF_ENEMIES + 1):
30             ts_intersection = ts[i[0]][1]
31
32             for j in i:
33                 if ts[j][1] is None:
34                     ts_intersection = None
35                     break
36
37             ts_intersection = ts_intersection.intersection(ts[j][1])
38             if len(ts_intersection) == 0:
39                 break
40
41         if ts_intersection is None:
42             continue
43
44         if len(ts_intersection) > 0:
45             t = ts_intersection.pop()
46             break
47
48     return t
```

---

---

## Листинг .9. Модуль BTS\_server

---

```
1 from secondary_functions import *
2 import argparse
3 from concurrent.futures import ThreadPoolExecutor
4 from collections import deque
5 from collections import Counter
6 import logging
7
8 TS = set()
9 RS = set()
10 QR = deque()
11
12 SERVER_ID = 0
13 SERVER_PORT = 0
14 SERVERS = []
15
16 DECISIONS1 = {}
17 AMOUNT_OF_DECISIONS1 = 0
18
19 DECISIONS2 = {}
20 AMOUNT_OF_DECISIONS2 = 0
21
22 THREAD_POOL_ON = True
23 IS_ACCEPTED1 = False
24 IS_ACCEPTED2 = False
25
26
27
28 def read_from_ts_file(file_name):
29     global TS
30
31     with open(file_name, 'r') as f:
32         data = json.load(f)
33         for i in data:
34             TS.add(tuple(i))
35
36
37 def match(temp, tup):
38     if len(temp) != len(tup):
39         return False
40
41     for i, j in zip(temp, tup):
42         if i is None:
43             continue
44         elif i != j:
45             return False
46
47     return True
```

---

---

**Листинг .10.** Модуль BTS\_server (продолжение)

---

```
1 def worker(client):
2     global THREAD_POOL_ON
3
4     req = get_message(client)
5
6     if req['op'] == 'out':
7         out(req['tup'])
8     elif req['op'] == 'rd':
9         rdp(client, req['temp'])
10    elif req['op'] == 'in':
11        inp(client, req['pid'], req['tup'])
12
13    elif req['op'] == 'enter':
14        enter_r(client, req['pid'], req['temp'])
15    elif req['op'] == 'exit':
16        exit_r(req['pid'])
17
18    elif req['op'] == 'accept1':
19        accept1(req['pid'], req['port'], req['tup'])
20    elif req['op'] == 'accept2':
21        accept2(req['pid'], req['port'], req['tup_dict'])
22
23    elif req['op'] == 'stop':
24        logging.info('WORKER: stop-message received')
25        THREAD_POOL_ON = False
26
27    else:
28        logging.info('WORKER: wrong request ' + str(req))
29
30    client.close()
31
32
33 def out(t):
34     global TS, RS
35
36     if t not in RS:
37         TS.add(t)
38         logging.info('OUT: add ' + str(t) + ' to TS')
39
40     try:
41         RS.remove(t)
42     except KeyError:
43         pass
44     else:
45         logging.info('OUT: remove ' + str(t) + ' from RS')
```

---

---

**Листинг .11.** Модуль BTS\_server (продолжение)

---

```
1 def rdp(sock, temp):
2     global TS
3     temp_set = set()
4
5     for t in TS:
6         if match(temp, t):
7             temp_set.add(t)
8
9     logging.info('RDP: temp_set for ' + str(temp) + ' : ' + str(
        temp_set))
10    if sock is not None:
11        send_message(sock, {'ts': temp_set})
12    else:
13        return temp_set
14
15    def inp(sock, pid, tup):
16        global TS, RS, QR
17        if QR[0][1] == pid:
18            d = paxos(sock, tup)
19            if d is not None:
20                if d not in TS:
21                    RS.add(d)
22                    logging.info('INP ' + str(pid) + ': add ' + str(d) + ' to
                        RS')
23                    TS.remove(d)
24
25    def enter_r(sock, pid, temp):
26        global QR
27        QR.append((sock, pid, temp))
28        if QR[0][1] == pid:
29            logging.info('ENTER_R: ' + str(pid) + ' enter cs')
30            temp_set = rdp(None, temp)
31            send_message(sock, {'resp': 'go', 'ts': temp_set})
```

---

---

**Листинг .12.** Модуль BTS\_server (продолжение)

---

```
1 parser = argparse.ArgumentParser()
2 parser.add_argument('id', type=int)
3 parser.add_argument('port', type=int)
4 parser.add_argument('TSFile', type=str)
5 parser.add_argument('quorum', nargs='+', type=int)
6 args = parser.parse_args()
7 SERVER_ID, SERVER_PORT, ts_file, SERVERS = args.id, args.port,
   args.TSFile, args.quorum
8
9 logging.basicConfig(filename=str(SERVER_ID) + 'log.txt', level=
   logging.DEBUG, format="%asctime)s - %(message)s")
10 logging.info('Server ' + str(SERVER_ID) + ' on port ' + str(
   SERVER_PORT))
11 read_from_ts_file(ts_file)
12 logging.info('Read from file: ' + str(ts_file))
13 s = socket(AF_INET, SOCK_STREAM)
14 s.bind('', SERVER_PORT)
15 s.listen(1)
16
17 with ThreadPoolExecutor(10) as pool:
18     while THREAD_POOL_ON:
19         try:
20             s.settimeout(30)
21             client_s, client_addr = s.accept()
22         except error:
23             pass
24         else:
25             pool.submit(worker, client_s)
26
27 s.close()
28 logging.info('RUN: stop working')
```

---