

10th International Conference "Distributed Computing and Grid
Technologies in Science and Education" (GRID'2023),
Dubna, Russia, July 3-7 2023

Distributed data transmission networks with a combined structure

Alexander M. Rappoport, Ilya I. Kurochkin

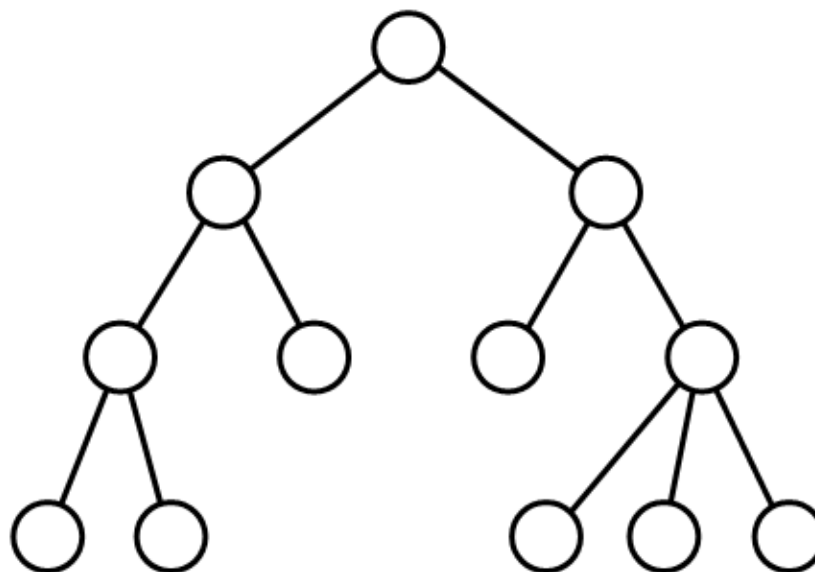
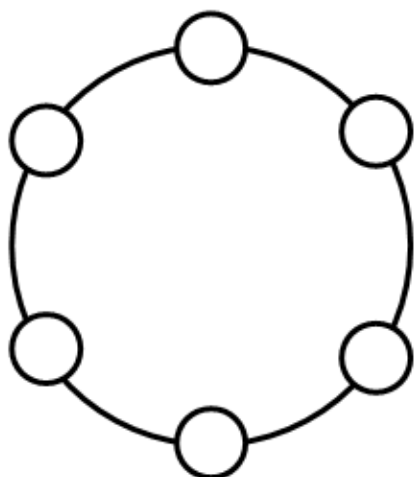
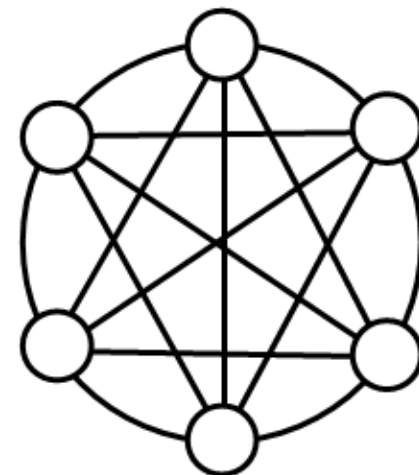
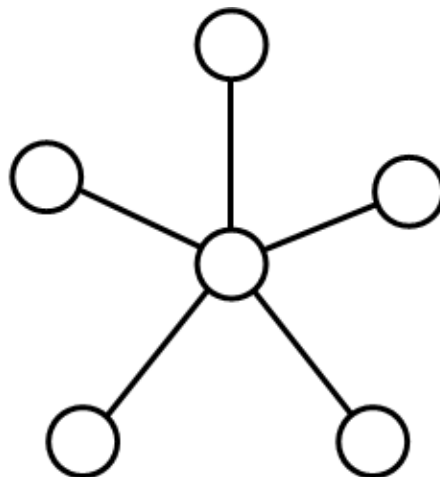
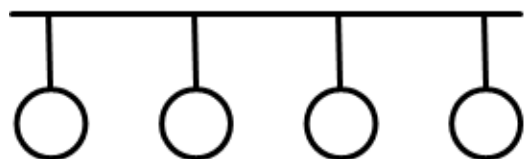
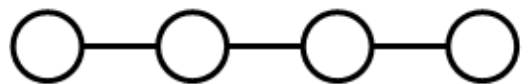
IITP RAS, Moscow, Russia

Распределенные сети коммуникаций
с комбинированной структурой

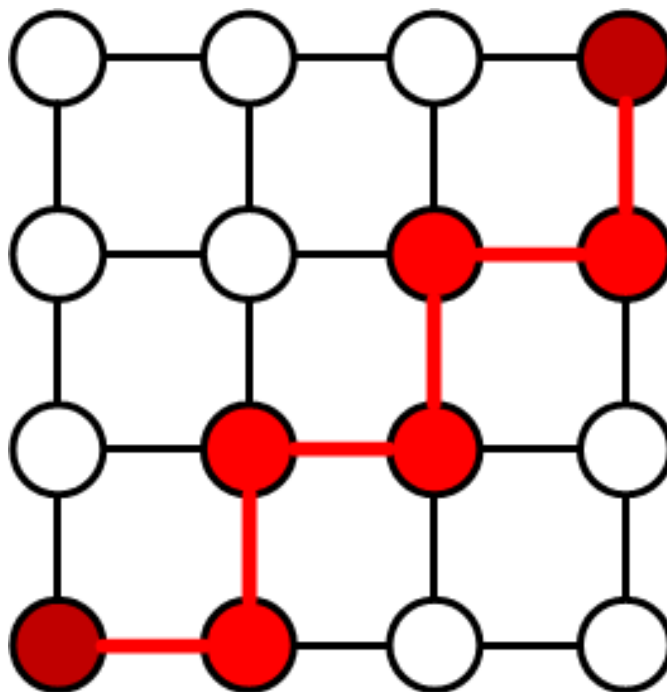
Раппопорт А.М., Курочкин И.И.

ЦРВ ИППИ РАН, Москва, Россия

Базовые сетевые топологии



Расстояние между узлами



Максимальное расстояние (диаметр)

Отказоустойчивость

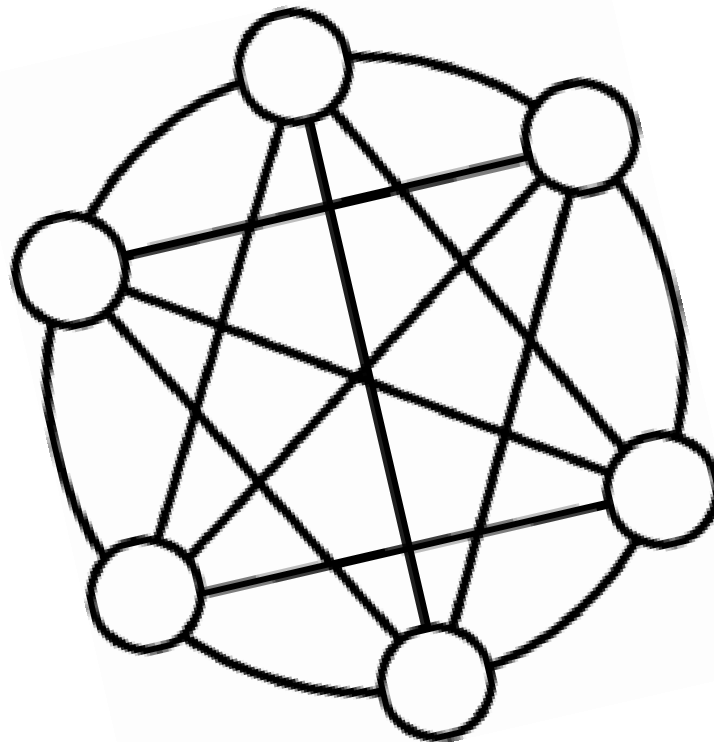
Проведение высокоэффективных распределенных вычислений естественно налагает определенные условия на **структуру взаимодействий** между участниками системы коммуникаций.

В случае единственного центрального элемента , осуществляющего управление и координацию, такой проблемы не существует, сеть является звездой и все остальные узлы достигаются за один шаг.

Значительно усложняется ситуация, когда имеется несколько «центральных» элементов и управляющие функции рассредоточены среди них, что способствует снижению уязвимости системы.

Возникает необходимость построения соответствующей конфигурации сети, позволяющей наилучшим способом организовать ее функционирование и обеспечить отказоустойчивость.

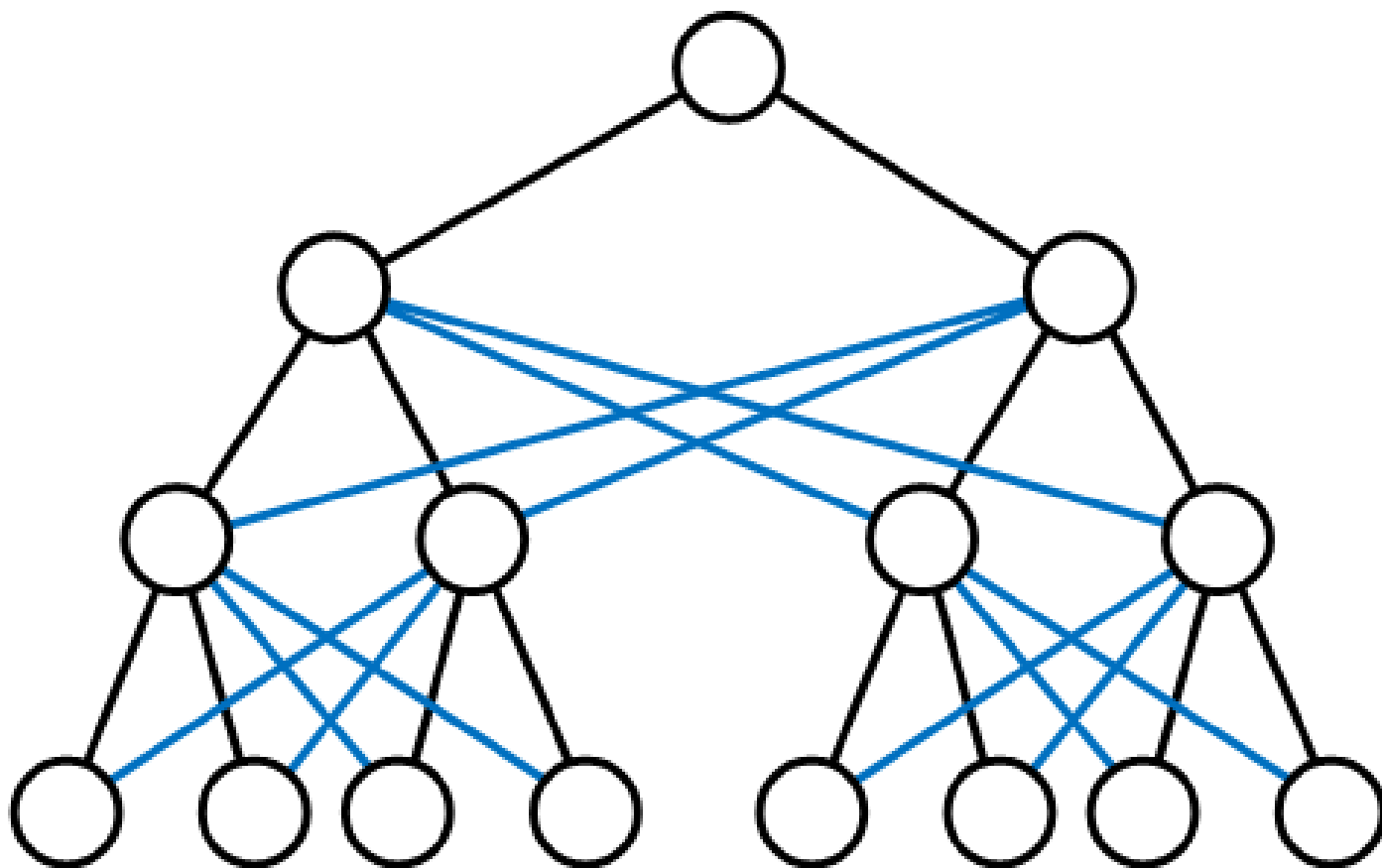
Полносвязная топология



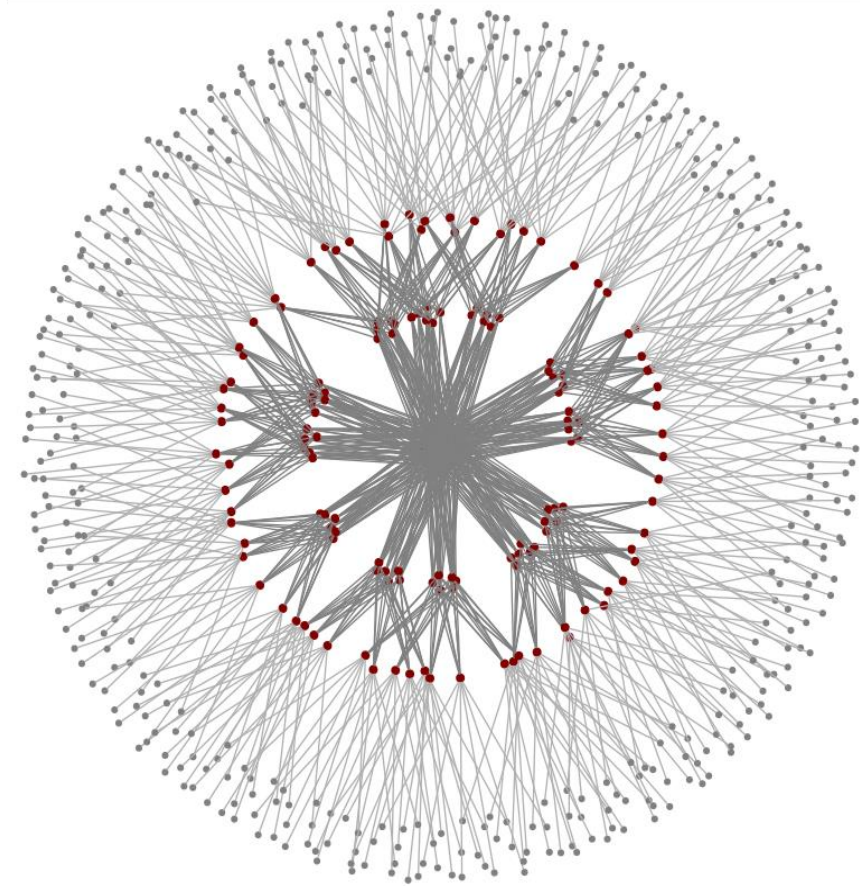
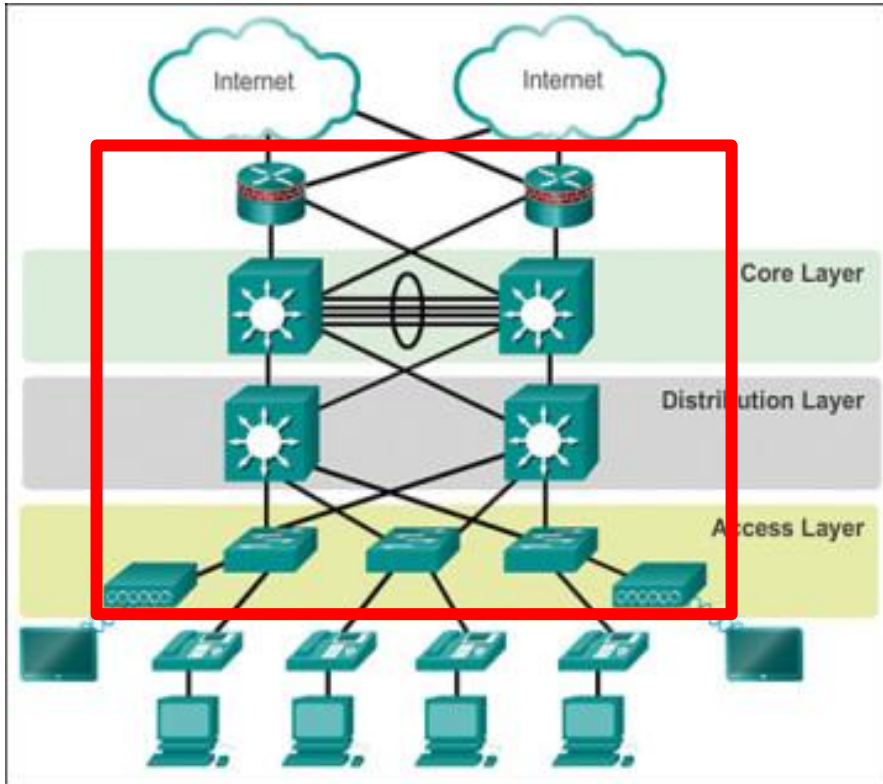
Сетевые топологии корпоративных сетей (ЦОД, суперкомпьютер и др.)

- **Fat tree**
- N-мерный тор
- N-мерная решетка
- Гиперкуб
- **Dragonfly**
- **Jellyfish**
- ...
- ~~Полносвязная сетевая топология~~

Fat tree



FatTree



Пусть $D = \{x_i, i = \overline{1, s}\}$ - некоторое доминирующее множество искомого графа G . В нем вершины из $D \subset V$ не смежны $(x_i, x_j) \notin E, i, j = \overline{1, s}$. Определим в $V \setminus D$ непересекающиеся «доминируемые» подмножества вершин $V_i = N(x_i)$ - окрестность вершины x_i и соответствующие им подграфы $G_i = G_i(V_i, E_i)$, по построению $V_i \cap V_j = \emptyset, E_i \cap E_j = \emptyset, i, j = \overline{1, s}$.

Потребуем выполнения дополнительных условий на структуру этих подграфов. И будем обозначать их (Y) совместно с уже сформулированными условиями. В этих подграфах вершины разбиты на 2 непересекающихся подмножества: $V_i = V_i^1 \cup V_i^2$, причем вершины из разных подмножеств не смежны. В подграфах $G_i^1 = G_i^1(V_i^1, E_i^1), G_i^2 = G_i^2(V_i^2, E_i^2)$ допустимы лишь внутренние ребра, а хотя бы одна вершина из V_i^2 смежна с вершиной из V_{i+1}^1 . Основным результатом состоял в следующем.

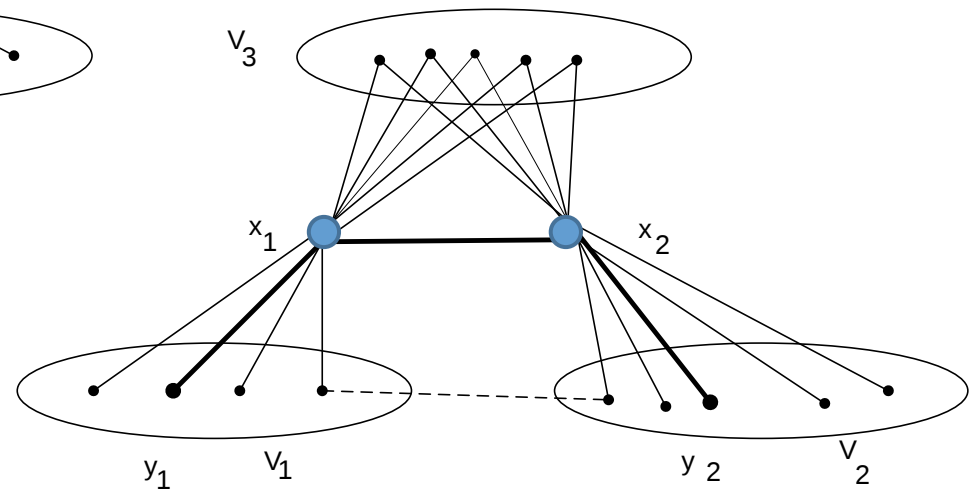
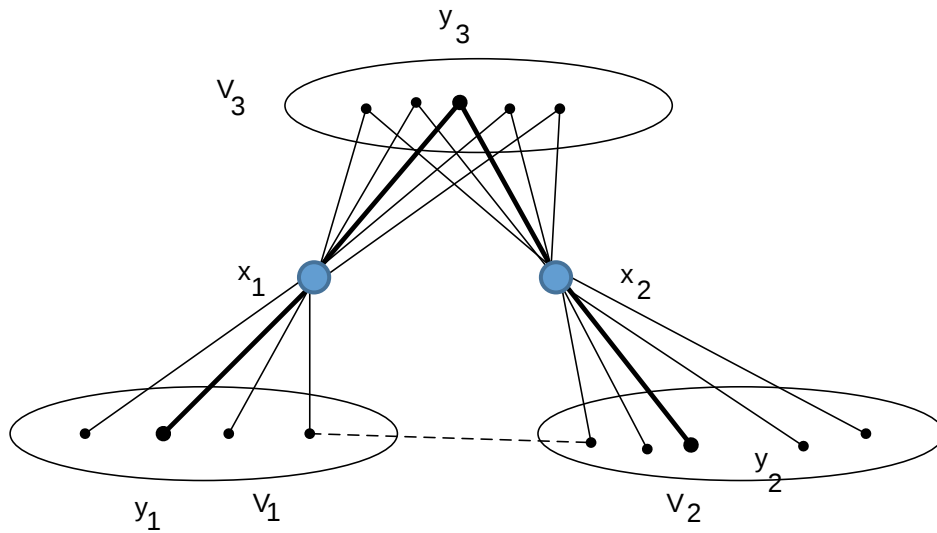
Теорема 1.

Если граф $G = G(V, E)$ с доминирующим множеством $D = \{x_i, i = \overline{1, s}\}$ удовлетворяет условиям (Y), то в нем длина максимальной кратчайшей цепи (диаметральной) равна $3s - 1$. Она имеет вид: $\{y_1^1, x_1, y_1^2, x_2, y_2^2, \dots, y_{s-1}^1, x_{s-1}, y_{s-1}^2, x_s, y_s^1\}$, где $y_i^1 \in V_i^1, y_i^2 \in V_i^2$. Кроме того, на совокупности всех графов с таким доминирующим множеством D диаметр $\delta \leq 3s - 1$.

Доминирующее множество

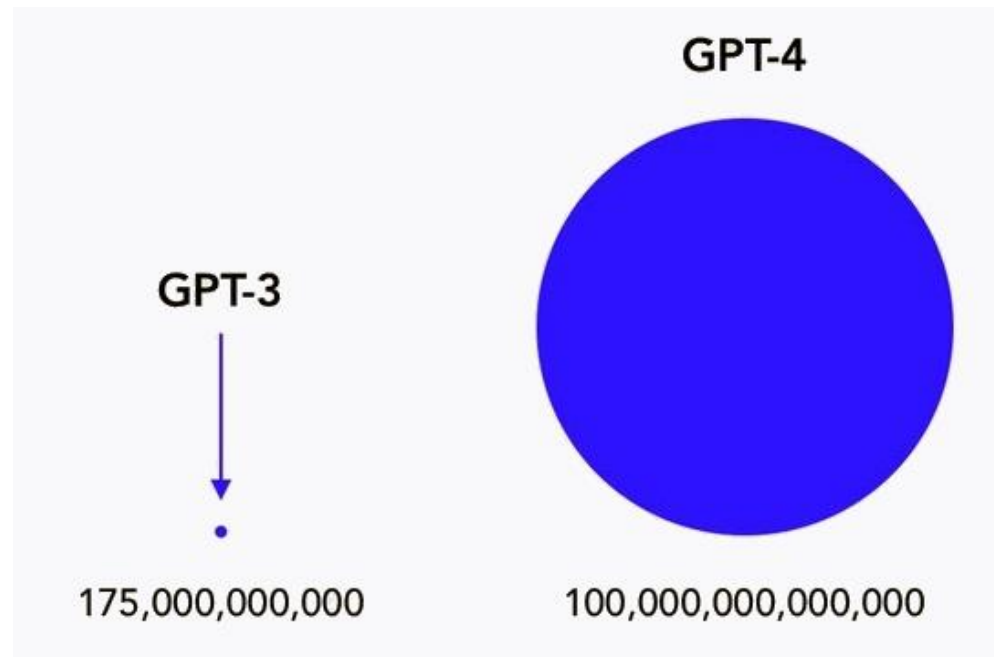
1. Нахождение доминирующего множества на существующем графе. Модификация графа – добавление ребер между вершинами доминирующего множества (прямая задача)
2. Изначально заданы только вершины графа и доминирующее множество вершин (всего 2 типа вершин). Нужно построить граф – добавить ребра с учетом ограничений существующего разбиения вершин графа на 2 типа (обратная задача)

Добавление ребер



Глубокие нейронные сети

- Тенденция к увеличению глубины нейронных сетей (152 у ResNet)
- Огромное количество настраиваемых параметров (ResNet-152 – 60.2 млн., GPT3 – 175 млрд.)



Датасет MNIST

MNIST - база данных рукописных цифр, содержащая 60 000 образов для обучения и 10 000 образов для тестирования. Размерность образов 28x28 пикселей.



Датасеты



ImageNet

The ImageNet dataset contains 14,197,122 annotated images according to the WordNet hierarchy. Since 2010 the dataset is used in the ImageNet Large Scale...

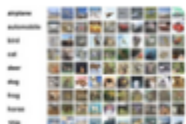
11,272 PAPERS • 106 BENCHMARKS



MNIST

The MNIST database (Modified National Institute of Standards and Technology database) is a large collection of handwritten digits. It has a training set of 60,0...

6,296 PAPERS • 50 BENCHMARKS



CIFAR-100

The CIFAR-100 dataset (Canadian Institute for Advanced Research, 100 classes) is a subset of the Tiny Images dataset and consists of 60000 32x32 color images....

6,076 PAPERS • 44 BENCHMARKS



SVHN (Street View House Numbers)

Street View House Numbers (SVHN) is a digit classification benchmark dataset that contains 600,000 32x32 RGB images of printed digits (from 0 to 9) cropped...

2,675 PAPERS • 11 BENCHMARKS



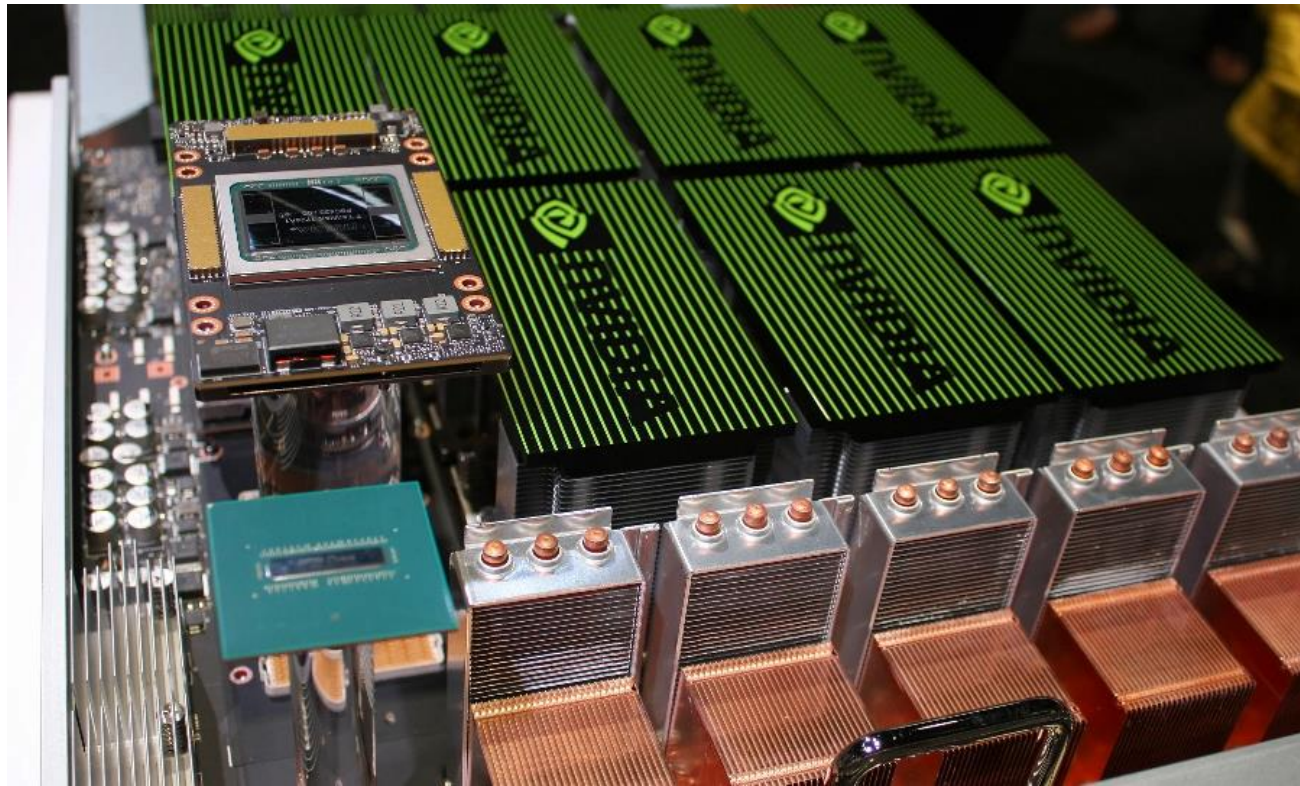
CelebA (CelebFaces Attributes Dataset)

CelebFaces Attributes dataset contains 202,599 face images of the size 178x218 from 10,177 celebrities, each annotated with 40 binary labels indicating facial a...

2,652 PAPERS • 18 BENCHMARKS

Узкие места обучения

- Большие размеры датасетов
- Необходимость задействовать весь датасет при обучении
- Время обучения нейронной сети



Виды распределенного обучения

Классификация 1:

- разделение по модели,
- разделение по данным.

Классификация 2 (по наличию параметрического сервера):

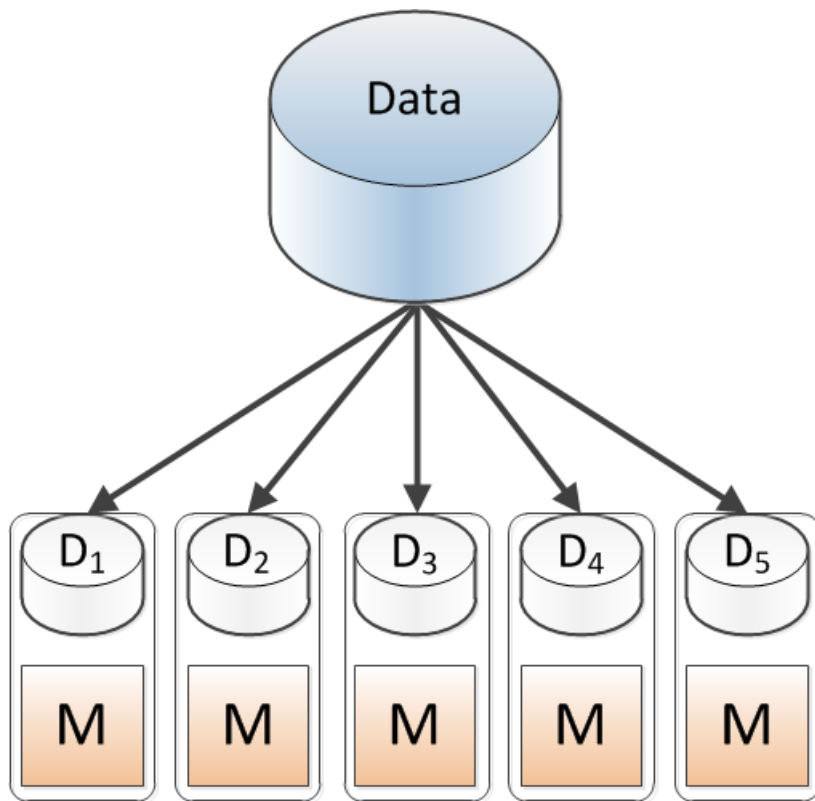
- с параметрическим сервером,
- **децентрализованное обучение.**

Классификация 3 (по ограничениям распространения данных):

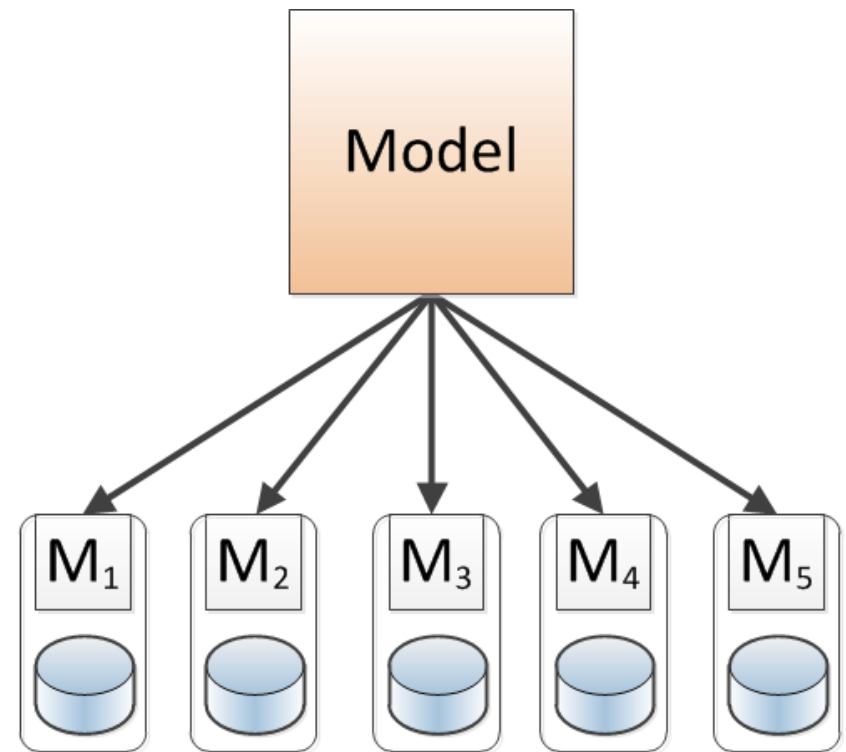
- с общими данными,
- федеративное обучение
 - Горизонтальное федеративное обучение;
 - Вертикальное федеративное обучение;
 - Федеративное «трансферное» обучение.

Распределенное обучение. Классификация 1

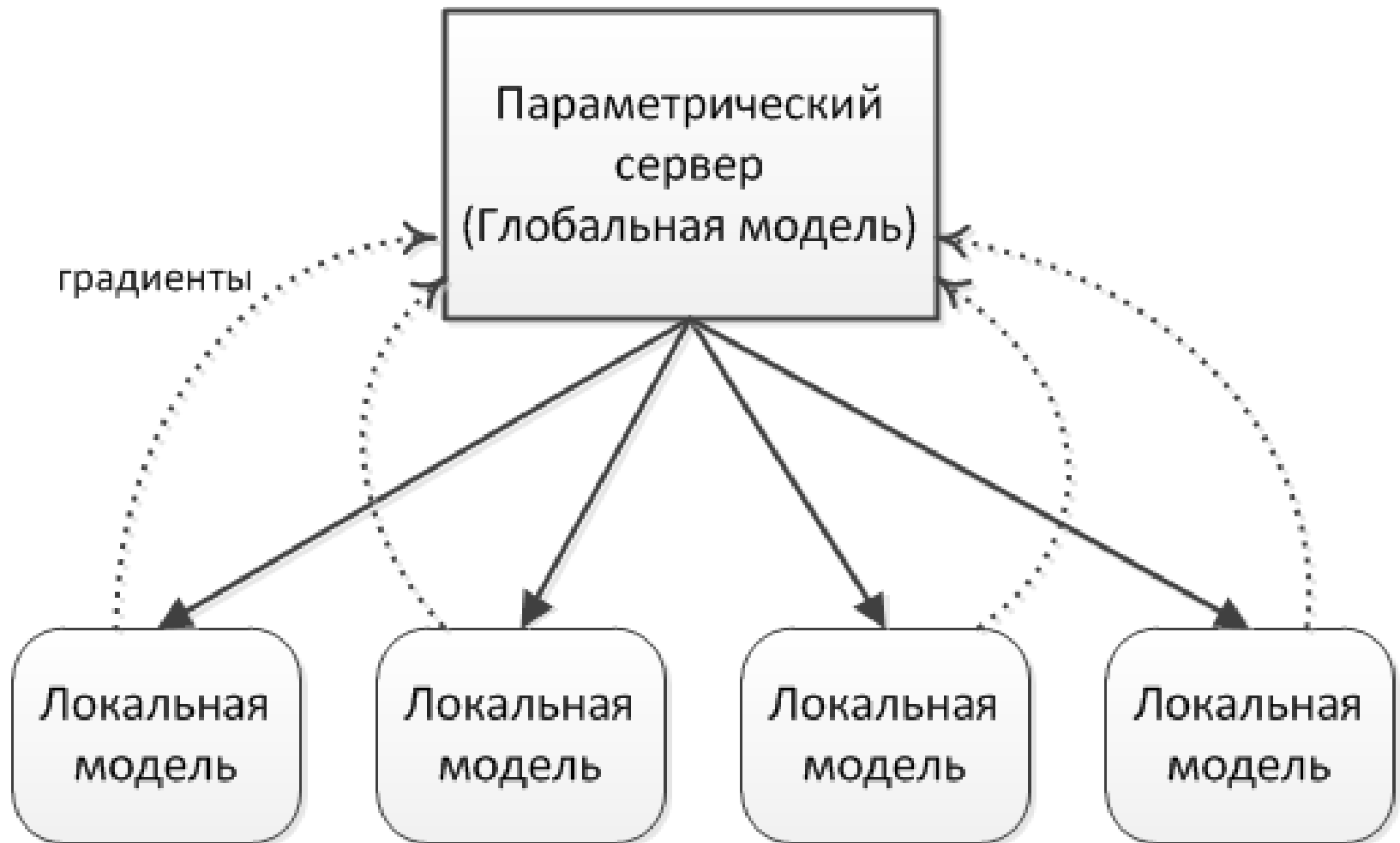
Разделение по данным



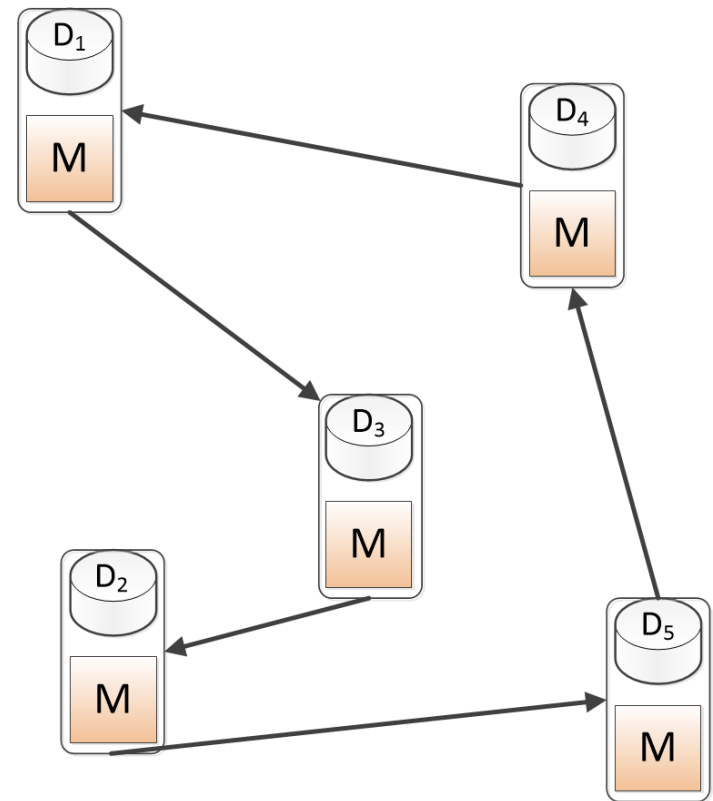
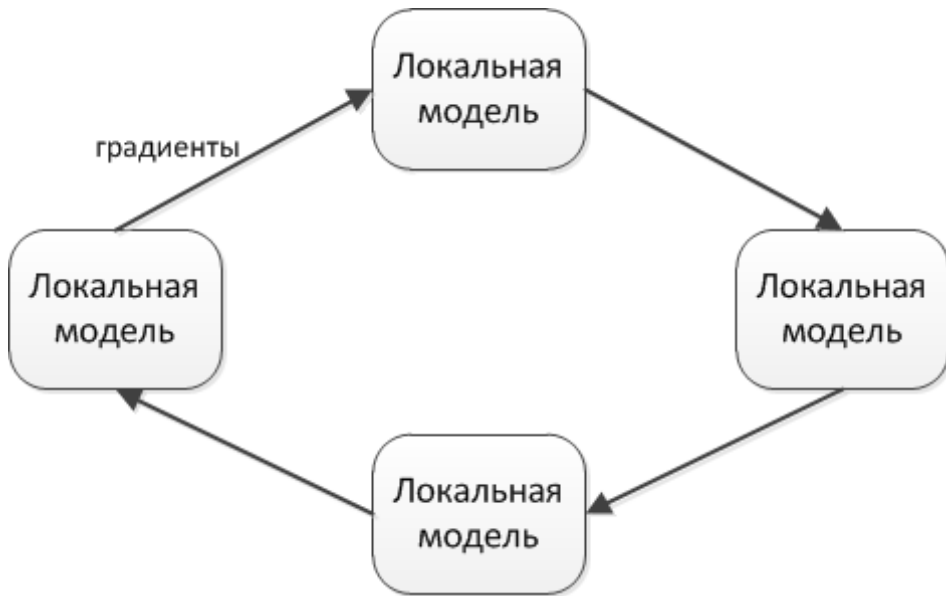
Разделение по модели



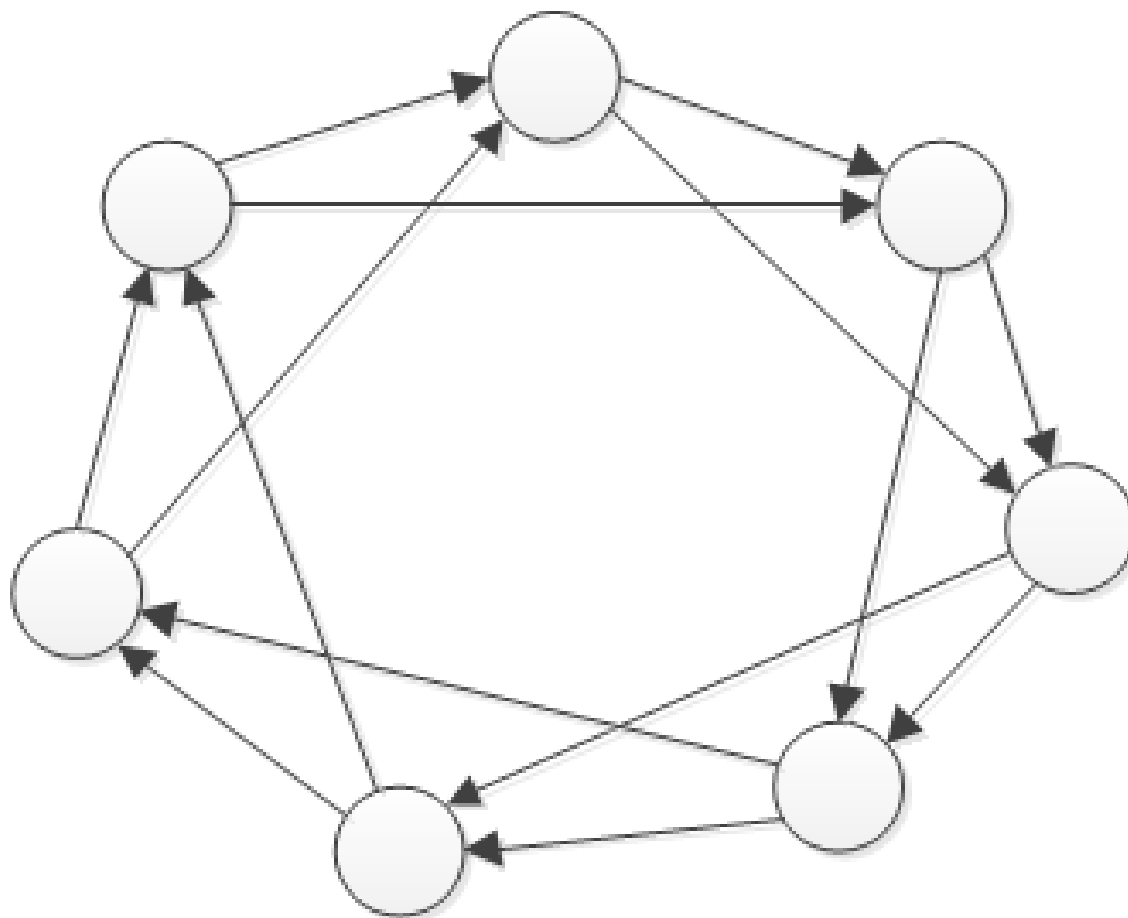
Распределенное обучение с параметрическим сервером



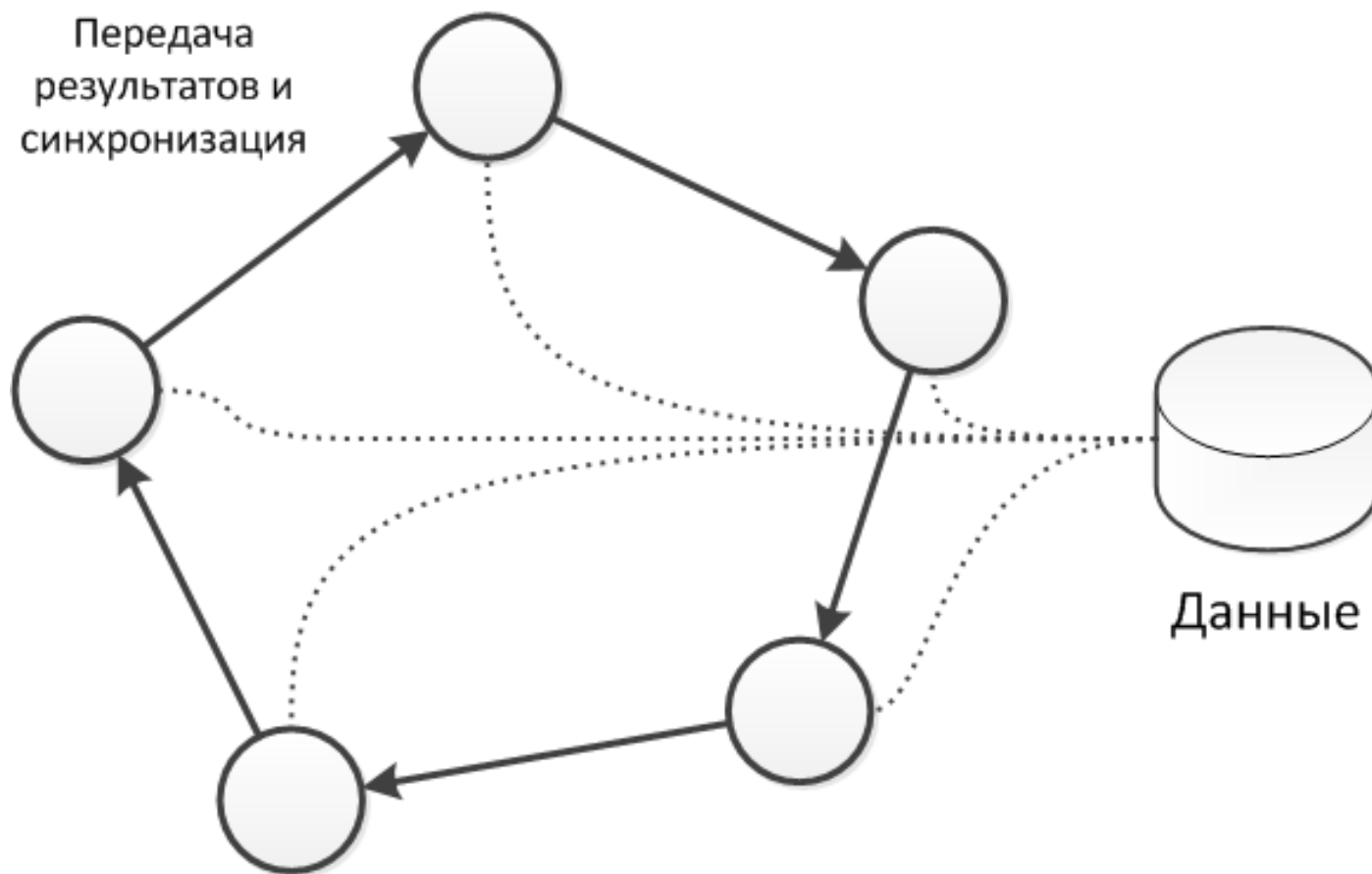
Децентрализованное глубокое обучение



Децентрализованное глубокое обучение

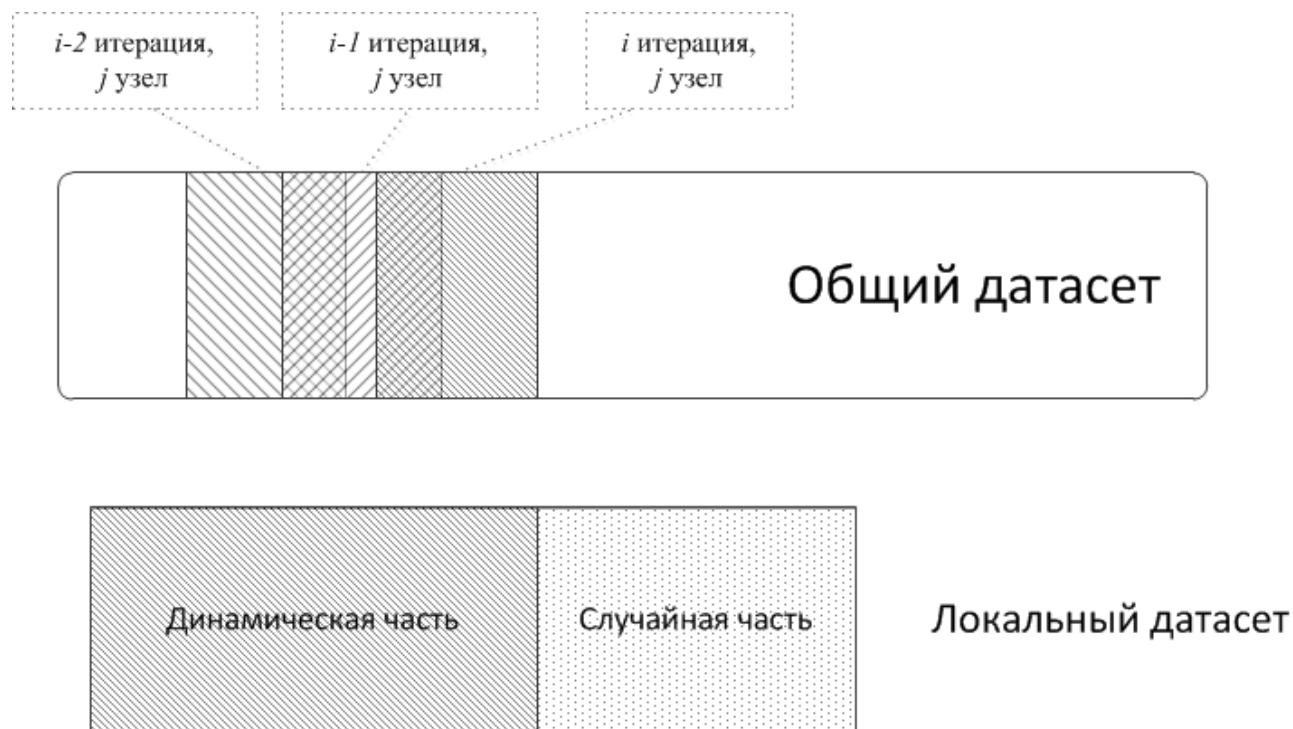


Децентрализованное глубокое обучение



Формирование локальных датасетов

- Стохастическое (случайное) формирование локального датасета;
- Разделение датасета на части (статическое) до конца обучения;
- Разделение датасета на части (динамическое) после каждой итерации;
- Гибридное формирование датасета с использованием, как случайного выбора, так наличия части изображений из локального датасета прошлой итерации.



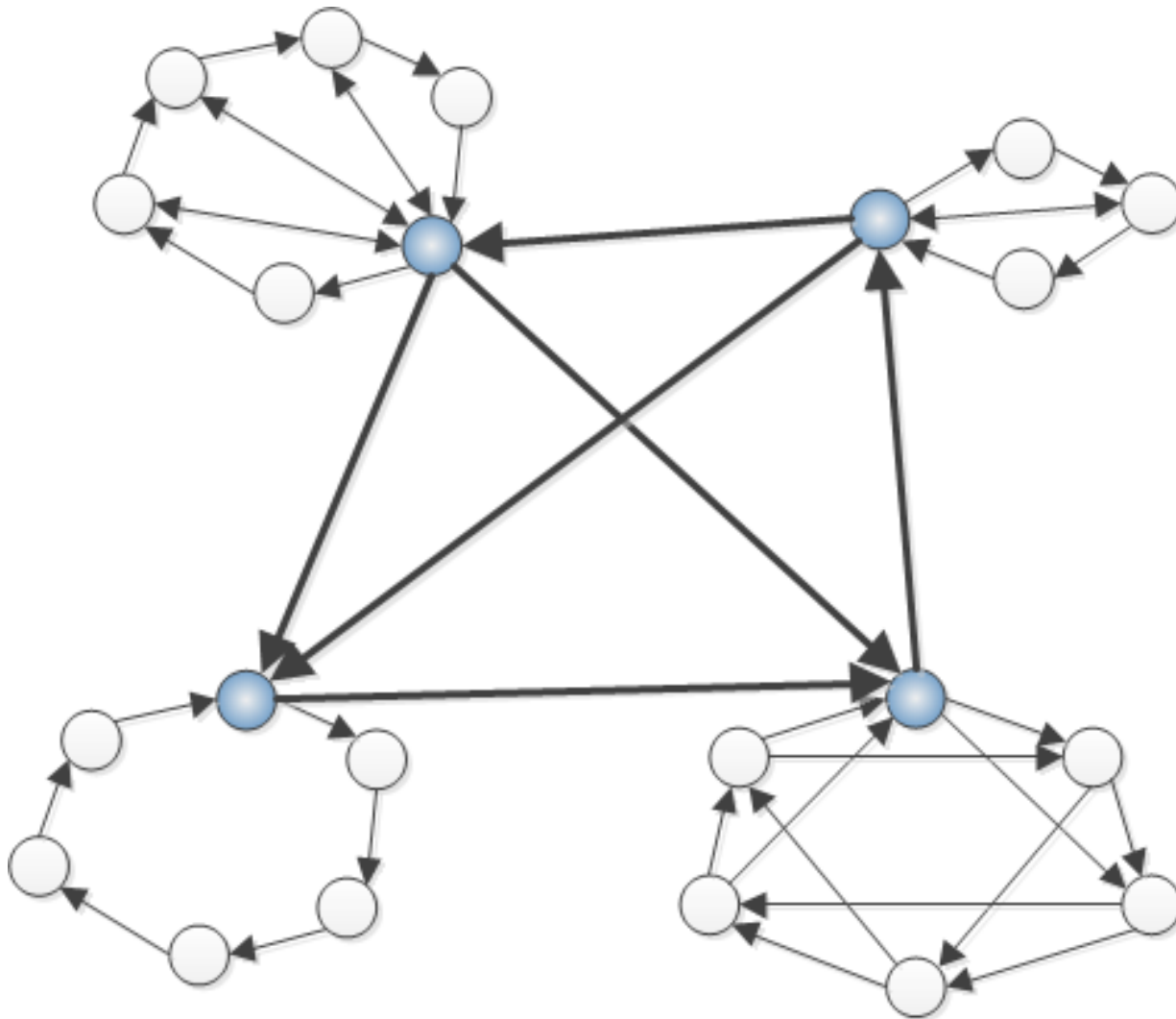
Особенности грид-систем

- Гетерогенность узлов распределенной системы;
- Автономность расчетов на различных узлах и невозможность постоянной координации расчетов между узлами;
- Ненадежность связей и возможное отключение вычислительных узлов;
- Непостоянное время непрерывной работы узла и трудность расчета длительных заданий;
- Наличие ошибок и задержек при расчетах.

Адаптация децентрализованного глубокого обучения для GRID-систем

- Выделение опорных/доверенных узлов в доминирующее множество
- Ускорение обмена градиентами в процессе глубокого обучения путем добавления ребер в графе

Использование модифицированных графов для организации децентрализованной передачи данных



Заключение

- Предложен метод построения графа с максимальным диаметром на совокупности графов с заданным размером доминирующего множества;
- Использование данного метода построения графа телекоммуникационной сети позволяет избежать чрезмерной централизации;
- Для представленных графов могут быть даны оценки (количества ребер / вершин / диаметра);
- Применение на задаче децентрализованного распределенного обучения.

Спасибо за внимание

**Центр распределенных вычислений
ИППИ РАН**

web: distributed-computing.ru

e-mail: kurochkin@iitp.ru