



Information Technologies  
and Mathematical Modelling  
named after A.F. Terpugov

2021  
December 1-5  
Tomsk, Russia  
online/offline

# ЗАДАЧА MARL ДЛЯ СЕТИ СВЕТОФОРОВ

Т.И. Тисленко, Д.В. Семенова  
Сибирский федеральный университет  
[tmtisko@mail.ru](mailto:tmtisko@mail.ru), [DVSemenova@sfu-kras.ru](mailto:DVSemenova@sfu-kras.ru)

## Актуальность

## Проблема

оптимизации планов координации светофорных объектов в условиях городской сети.



**Рисунок 1:** Пробки в Красноярске

# Существующие модели адаптивных систем светофоров

Таблица 1: Модели адаптивных систем светофоров

Критерий	UTCS-1	SCOOT	OPAC	MARL
город	Вашингтон	Лондон	Арлингтон, Тускон	Торонто
временной период	1970е	1995	1983, 1989	2010
длительность фаз	фиксированная			переменная
оптимизация	оффлайн			онлайн
предсказание	нет	есть		нет
устройство	централизованная			децентрализованная
основные ограничения	постоянный сбор данных	сенсоры далеко	только для 8 фаз	«ПРОКЛЯТИЕ РАЗМЕРНОСТИ»

# Цели и задачи

## Цель работы

Разработка и исследование математической модели мультиагентной системы для задачи координации светофорных объектов в транспортной сети мегаполиса.

## Задачи

1. Построить математическую модель MARL для сети двухфазных светофоров.
2. Разработать алгоритм решения задачи MARL.
3. Провести вычислительные эксперименты для одного и двух светофоров.

# Основные определения и обозначения

Таблица 2: Основные обозначения

$K$	число агентов
$k$	номер агента
$S^k$	пространство состояний для $k$ -го агента
$A^k$	множество решений для $k$ -го агента
$s_t$	совокупное состояние среды в момент времени $t$
$a_t$	совокупное управление в момент времени $t$
$r(s_t, a_t; s_{t+1})$	время, подсчитанное для машин двигающихся на фазе $s_{t+1}$
$N(k)$	множество соседних с $k$ агентов
$s^{kj}$	совместные состояния агентов $k$ и $j$
$a^{kj}$	совместные действия агентов $k$ и $j$
$V(\{S_t, \delta\})$	функция суммарных доходов/вознаграждений для $K$ агентов

# Основные определения и обозначения

## Для $k$ -го агента

- $S^k$  определяются классом вычетов по модулю

$$S^k = n^{(k)}\mathbb{Z} = \left\{ s^{(0)} = 0, s^{(1)} = 1, \dots, s^{(n^{(k)}-1)} = n^{(k)} - 1 \right\}, \quad n^{(k)} = |S^k|,$$

где  $s^{(i)}$  интерпретируется как «активна фаза  $i$ »;

- $A^k$  определяется характеристикой кольца  $n^{(k)}\mathbb{Z}$ , т.е. не превышает ее,  $A^k = \left\{ a^{(0)} = 0, a^{(1)} = 1, \dots, a^{(n^{(k)}-1)} = n^{(k)} - 1 \right\}$ , где  $a^{(j)}$  интерпретируется как «сменить фазу  $j$  раз».

## В момент времени $t$

- $\mathbf{s}_t = \{s_t^1, s_t^2, \dots, s_t^K\} \in \mathcal{S} = S^1 \times S^2 \times \dots \times S^K$ ;
- $\mathbf{a}_t = \{a_t^1, a_t^2, \dots, a_t^K\} \in \mathcal{A} = A^1 \times A^2 \times \dots \times A^K$ ;
- $\mathbf{a}^{kj} \in A^k \times A^j$  и  $\mathbf{s}^{kj} \in S^k \times S^j$ ;
- $V(\{\mathbf{s}_t, \delta\}) = \sum_{t=0}^{\infty} \gamma^t r(\mathbf{s}_t, \mathbf{a}_t)$ , где  $0 \leq \gamma \leq 1$  — коэффициент переоценки,  $\{\mathbf{s}_t\}$  — реализация случайного процесса при выбранной стратегии  $\delta = \{\mathbf{a}_t, 0 \leq t < \infty\}$ .

# Постановка задачи MARL

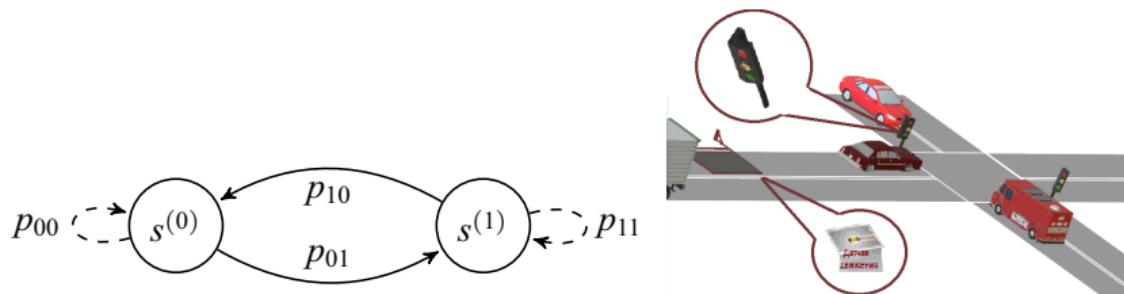


Рисунок 2: Стохастический граф управляемого процесса смены фаз для двухфазного светофора и его визуальная интерпретация. Пунктиром обозначено действие  $a^{(0)}$ , а сплошной —  $a^{(1)}$

# Постановка задачи

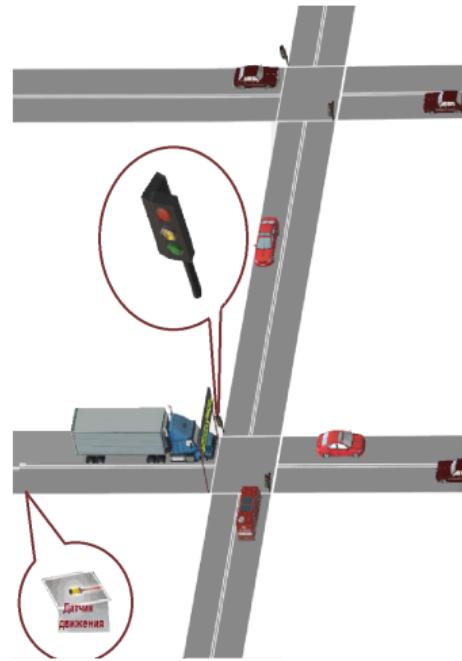
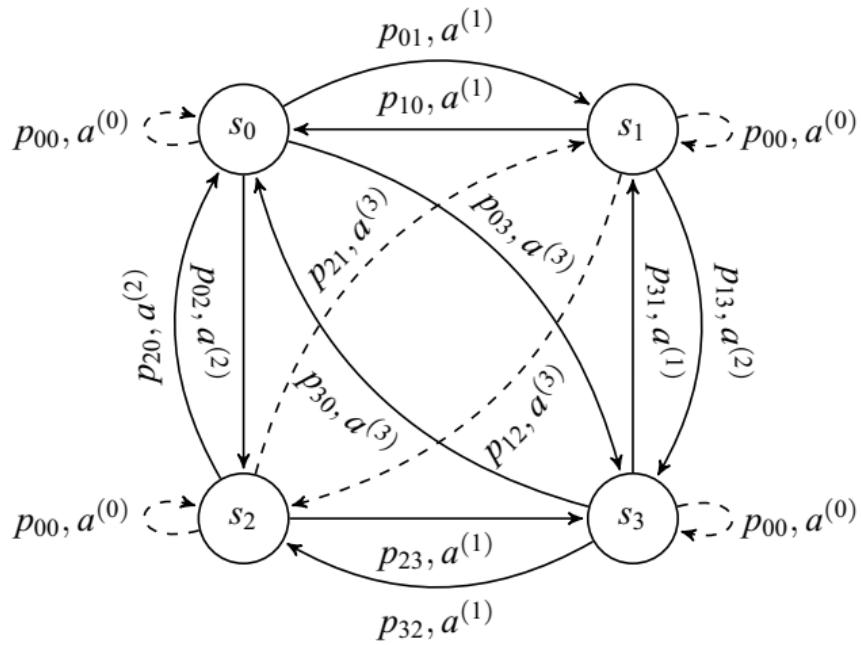


Рисунок 3: Стохастический граф сети из двух двухфазных светофоров

# Задача MARL для сети светофоров

## Постановка задачи MARL для сети светофоров

Требуется найти такое управление  $\delta = \{\mathbf{a}_t, 0 \leq t < \infty\}$ , которое доставляет максимум функции суммарных вознаграждений для  $K$  агентов

$$V(\{\mathbf{s}_t, \delta\}) = \sum_{t=0}^{\infty} \gamma^t r(\mathbf{s}_t, \mathbf{a}_t) \rightarrow \max,$$

где  $0 \leq \gamma \leq 1$  — коэффициент переоценки,  $\{\mathbf{s}_t\}$  — реализация случайного процесса.

Решение ищется методом динамического программирования на основе принципа оптимальности Беллмана. Функцию суммарных вознаграждений при оптимальном управлении на шаге  $t$ :

$$V^*(\{\mathbf{s}_{t'}\}_{t'=0}^{t'=t}) = \max_{\mathbf{a} \in \mathcal{A}} Q_t(\mathbf{s}_t, \mathbf{a}), \quad (1)$$

где

$$Q_t(\mathbf{s}_t, \mathbf{a}) = \sum_{\mathbf{s}_{t+1} \in \mathcal{S}} p(\mathbf{s}_t, \mathbf{a}; \mathbf{s}_{t+1}) \left( r(\mathbf{s}_t, \mathbf{a}; \mathbf{s}_{t+1}) + \gamma \max_{\mathbf{a}' \in \mathcal{A}} Q_{t-1}(\mathbf{s}_{t+1}, \mathbf{a}') \right). \quad (2)$$

# Решение задачи MARL для двух светофоров

- Множество соседних агентов  $N(k) = \{j\}$ ,  $k \neq j$ ,  $k, j = 0, 1$ .
- Функция обучения агента  $k$

$$Q_t^k(\mathbf{s}_t, a_t^k) = \sum_{a_j \in A^j} \underbrace{p(a_t^k; \mathbf{a}_t^{kj}) Q_t(\mathbf{s}_t, \mathbf{a}_t^{kj})}_{Q_t^{kj}(\mathbf{s}_t, \mathbf{a}_t^{kj})} = \sum_{a^j \in A^j} Q_t^{kj}(\mathbf{s}, \mathbf{a}_t^{kj}). \quad (3)$$

$$\begin{aligned} Q_t^k(\mathbf{s}_t, a_t^k) &= \sum_{a_j \in A^j} p(a_t^k; \mathbf{a}_t^{kj}) \left( \sum_{\mathbf{s}_{t+1} \in \mathcal{S}} p(\mathbf{s}_t, \mathbf{a}_t^{kj}; \mathbf{s}_{t+1}) \left( r(\mathbf{s}_t, \mathbf{a}_t^{kj}; \mathbf{s}_{t+1}) + \gamma \max_{\mathbf{a}' \in \mathcal{A}} Q_{t-1}(\mathbf{s}_{t+1}, \mathbf{a}') \right) \right) = \\ &= \sum_{\mathbf{s}_{t+1} \in \mathcal{S}} \sum_{a_j \in A^j} p(\mathbf{s}_t, a_t^k; \mathbf{s}_{t+1}) \left( r(\mathbf{s}_t, \mathbf{a}_t^{kj}; \mathbf{s}_{t+1}) + \gamma \max_{\mathbf{a}' \in \mathcal{A}} Q_{t-1}(\mathbf{s}_{t+1}, \mathbf{a}') \right). \end{aligned} \quad (4)$$

Оптимальное решение для агента  $k$  на шаге  $t$

$$a_t^k = \arg \max_{a^k \in A^k} \sum_{a^j \in A^j} Q_t^{kj}(\mathbf{s}, \mathbf{a}^{kj}). \quad (5)$$

# Решение задачи MARL для сети светофоров

Имея алгоритм решения задачи для двух светофоров, можно получить решение для любого их количества в сети.

## Алгоритм MARLIN: основная идея

Оптимальное управление для фиксированного агента  $k$  будем искать как решение задачи MARL в виде:

$$a_t^k = \arg \max_{a^k \in A^k} \sum_{j \in N(k)} \sum_{a^j \in A^j} Q_t^{kj}(\mathbf{s}, \mathbf{a}^{kj}) p(\mathbf{s}' | (\mathbf{s}, \mathbf{a}^{kj})) \quad (6)$$

El-Tantawy S., Abdulhai B. and Abdelgawad H., Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC) // IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 3. 2013. — Р. 1140-1150.

# Вычислительные эксперименты

Вычислительные эксперименты проводились в системе имитационного моделирования Anylogic. Алгоритмы реализованы на языке программирования Java 8. Построены модели перекрестка с фиксированной длительностью фаз (рис. 4а) и управляемого марковским процессом (рис. 4б).

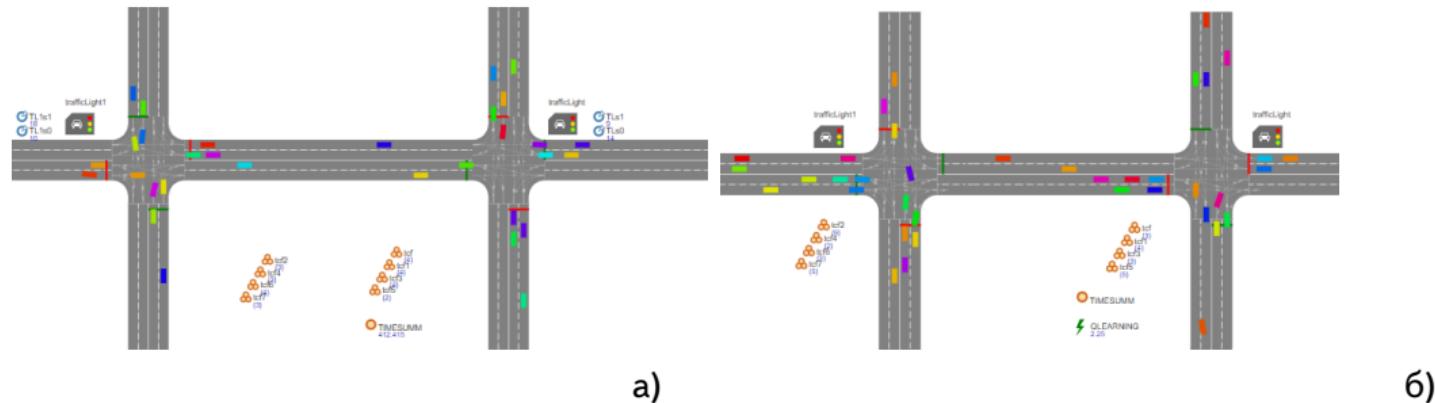


Рисунок 4: а) модель перекрестка с фиксированной длительностью фаз, б) модель перекрестка управляемого марковским процессом

# Основные результаты работы

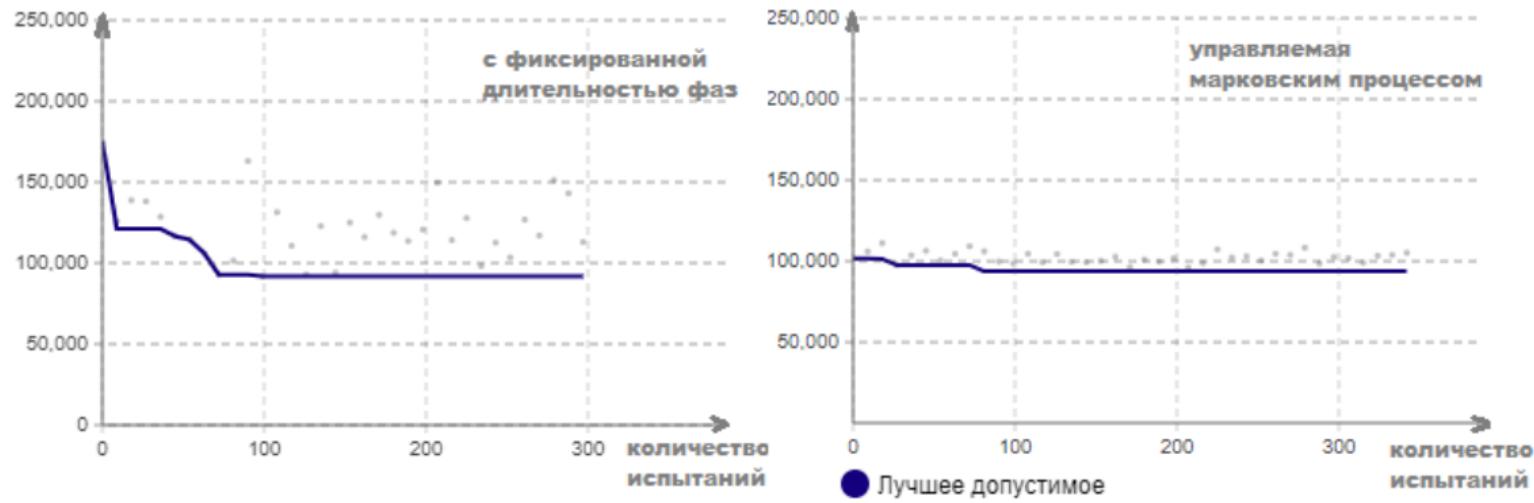


Рисунок 5: Сравнение задержки *TIMESUMM*

# Основные результаты

- Построена математическая модель процесса выбора фазы для сети светофоров, отличающаяся учетом текущего расположения светофоров и их загрузки и позволяющая сформулировать оптимизационные задачи, целью которых является минимизация задержки трафика автомобилей.
- Разработана структура мультиагентной системы — сети сферетофоров участка дороги.
- Проведены вычислительные эксперименты в системе имитационного моделирования Anylogic для одного и двух сферетофоров.

# Перспективы исследования

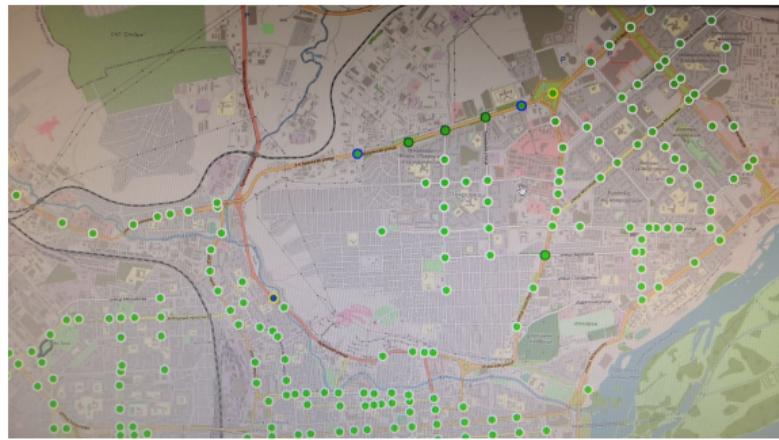


Рисунок 6: Карта перекрестков, предоставленная УдиБ г. Красноярска

# Перспективы исследования



Рисунок 7: Модель перекрестка проспект Свободный – ул. Годенко

# Основная литература

-  El-Tantawy S., Abdulhai B. and Abdeltawab H., Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC) // IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 3. 2013. — P. 1140-1150.
-  Лекции по случайным процессам: учебное пособие / А. В. Гасников, Э. А. Горбунов, С. А. Гуз и др.; под ред. А. В. Гасникова. – «Москва»: МФТИ, 2019. — 285 с.
-  Марковские процессы принятия решений. / Майн Х., Осаки С. Главная редакция физико-математической литературы издательства «Наука», 1977. — 176 с.
-  Sandholm T.W. Contract Types for Satisficing Task Allocation: I Theoretical Results // AAAI Spring Symposium Series: Satisficing Models. 1998. — P. 68-75.
-  Управляемые марковские процессы с конечными пространствами состояний и управлений, Теория вероятностей и ее примен., Том 11. /В. В. Рыков. 1966. — 343-351 с.



Information Technologies  
and Mathematical Modelling  
named after A.F. Terpugov

**2021**  
December 1-5  
Tomsk, Russia  
online/offline

**Спасибо за внимание!**