МАСКИРОВАНИЕ ТОПОЛОГИЧЕСКИХ СВОЙСТВ ВЫЧИСЛИТЕЛЬНЫХ СЕТЕЙ В УСЛОВИЯХ СЕТЕВОЙ РАЗВЕДКИ. Часть 2

Горбачев А. А.[1]

DOI: 10.21681/2311-3456-2025-1-63-72

Цель исследования: *разработка системы моделей, включающей классические модели случайных графов и генеративные модели искусственного интеллекта и предназначенной для решения задачи маскирования топологических свойств вычислительных сетей при генерации ложного сетевого трафика и применении ложных сетевых информационных объектов, позволяющих, с одной стороны, обеспечить заданную степень сходства топологических свойств реальных вычислительных сетей с ложными, а с другой стороны, максимизировать показатель защищенности критических узлов реальных вычислительных сетей.*

Используемые методы*: взвешенная аддитивная линейная свертка, случайный граф Эрдеша-Реньи, Барбаши, Ваттса-Строгаца, Харари, алгоритм байесовской оптимизации, модель сверточного вариационного автокодировщика, модель графового вариационного автокодировщика.*

Результат исследования: *представленная система моделей позволяет повысить результативность защиты вычислительной сети за счет формирования у злоумышленника устойчивого ложного представления относительно топологических свойств вычислительной сети с учетом повышения защищенности критических узлов посредством смещения положения ложных критических узлов по отношению к реальным, при обеспечении заданной степени сходства ложной топологии вычислительной сети по отношению к реальной топологии. Система моделей включает в себя конвейер машинного обучения на основе моделей случайных графов Эрдеша-Реньи, Барбаши, Ваттса-Строгаца, Харари, используемых для формирования обучающего набора данных, модели графового вариационного автокодировщика, модели выборки из скрытого пространства, содержащей показатели качества генерируемой ложной структуры, эволюционного алгоритма скалярной оптимизации, осуществляющего поиск оптимальной точки синтеза ложной структуры в скрытом пространстве вариационного автокодировщика, а также генератор ложного трафика, реализующего отправку пакетов с заданными сетевыми идентификаторами. Разработанный конвейер имеет ограничения по размерности синтезируемой ложной топологии в связи с вычислительной сложностью процесса обучения генеративной модели и поиска оптимальной точки синтеза.*

Научная новизна: *заключается в применении байесовского алгоритма оптимизации для выбора оптимальной точки синтеза ложной топологии из скрытого пространства обученного графового вариационного автокодировщика, в использовании целевой функции, представленной линейной взвешенной сверткой из коэффициента Жаккара между множеством ребер ложной и реальной топологии вычислительной сети, показателей защищенности вычислительной сети: среднего кратчайшего расстояния между реальными и ложными критическими узлами, коэффициента Жаккара между множеством ложных и реальных критических узлов вычислительной сети. В применении моделей случайных графов для формирования обучающего набора данных.*

Ключевые слова: *ложные информационные объекты, вариационный автокодировщик, конвейер машинного обучения, искусственный интеллект, оптимизация, метаэвристические алгоритмы, случайные графы.*

Введение

Одним из методов снижения эффективности анализа сетевого трафика злоумышленниками является генерация ложного сетевого трафика с сетевыми параметрами ложной вычислительной сети (идентификаторами сетевого, транспортного или прикладного уровня, динамическими характеристиками, структурными атрибутами) [1–4]. Как это было показано в первой части настоящей работы синтез ложных топологий вычислительных сетей за счет решения задачи комбинаторной оптимизации матрицы смежности графа с ростом количества вершин графа становится вычислительно неэффективным, поэтому для решения задачи генерации сетей различной природы используются методы снижения размерности, вложения графа (*эмбеддинга графа, graph embedding*) в пространство меньшей размерности, параметрические модели случайных графов[2,3], а также эвристические алгоритмы[4] [5]. При синтезе сложных сетей,

1 Горбачев Александр Александрович, кандидат технических наук, Краснодарское высшее военное орденов Жукова и Октябрьской Революции Краснознаменное училище имени генерала армии С. М. Штеменко, г. Краснодар, Россия. E-mail: infosec23.00@mail.ru

2 Haddadi H. Topological Characteristics of IP Networks. University College London. 2008. 114 p.

3 Liu W., Chen P., Yu F., Suzumura T., Hu G. Learning Graph Topological Features via GAN. IEEE. 2019. Vol. 7. pp. 21834–21843.

4 Батенков К. А. Анализ и синтез структур сетей связи методом перебора состояний // Вестник Санкт-Петербургского университета. Прикладная математика. Процессы управления. 2022. Т. 18. Вып. 3. С. 300–315.

Рис. 1. Классификация генеративных моделей машинного обучения

учитывающих веса и атрибуты вершин графов, используют вложения графов с использованием методов случайного блуждания, матричного разложения, алгоритмов глубокого обучения.

Представления многомерных входных данных о структуре сложных сетей осуществляется с целью кластеризации и визуализации непараметрическими методами (k-ближайших соседей, k-средних, методом главных компонент, стохастическим вложением соседей с t-распределением и др.) либо для параметризации моделей случайных графов (стохастических Кронекеровских графов, модели Ваксмана[5], экспоненциальных моделей случайных графов, безразмерных моделей случайных графов и др.), моделей глубокого обучения [6–8]. Параметрические генеративные модели используются с целью генерации графов, прогнозирования связей (ребер), классификации, кластеризации вершин, подграфов, графов [9–12]. В контексте защиты критических узлов вычислительных сетей необходимо, чтобы модель генерировала ложные структуры не только сходные с конкретной топологией (или совокупностью топологий) вычислительной сети, но и позволяла генерировать топологии в *управляемом непрерывном диапазоне* топологических характеристик со смещенным положением ложных критических узлов относительно реальных критических узлов.

С другой стороны, генерация данных произвольной природы из аппроксимированного распределения в настоящее время успешно реализуется моделями, методами и алгоритмами генеративного искусственного интеллекта (генерация текста, музыки, изображений) [13]. Генеративные модели искусственного интеллекта используют различные подходы к моделированию истинного распределения наблюдений и представлены широким классом моделей глубокого обучения, в большинстве случаев, основанных на архитектурах искусственных нейронных сетей (рис. 1).

В соответствии с теоремой «*об отсутствии бесплатных завтраков*»[6] не существует возможности априорного выбора наилучшей универсальной модели или алгоритма оптимизации для решения конкретной задачи, но полный перебор существующих моделей и алгоритмов является практически нецелесообразным, поэтому редукция многообразия методов и моделей генеративного искусственного интеллекта осуществляется на основе критериев, представленных ниже.

Использование алгоритмов машинного обучения без учителя. Функционирование вычислительных сетей связано с обработкой неструктурированных данных и постоянный анализ сетевого трафика не предполагает заблаговременного создания размеченных данных. Необходимо обеспечить возможность оперативного переобучения без проведения предварительной разметки данных.

Возможность синтеза топологии с заданными характеристиками. Модель должна позволять синтезировать структуры ложных вычислительных сетей, которые обладают заданными характеристиками качества с незначительным отклонением. В связи с чем, целесообразно использовать алгоритмы, моделирующие распределение топологических характеристик в явной форме.

Высокая вариабельность характеристик синтезируемых топологий. Генеративные модели должны обеспечивать синтез топологий как полностью идентичных исходным топологиям, так и промежуточных вариантов с заданными характеристиками. Данное требование связано с непрерывностью распределения характеристик синтезируемых топологий в скрытом пространстве моделей.

5 Waxman Bernard M. Routing of multipoint connections. IEEE journal on selected areas in communications. 1988. Vol. 6, no. 9. pp. 1617–1622.

6 Wolpert D.H., Macready W.G. No free lunch theorems for optimization. IEEE transactions on evolutionary computation. 1997. Vol №. 1. pp. 67–82.

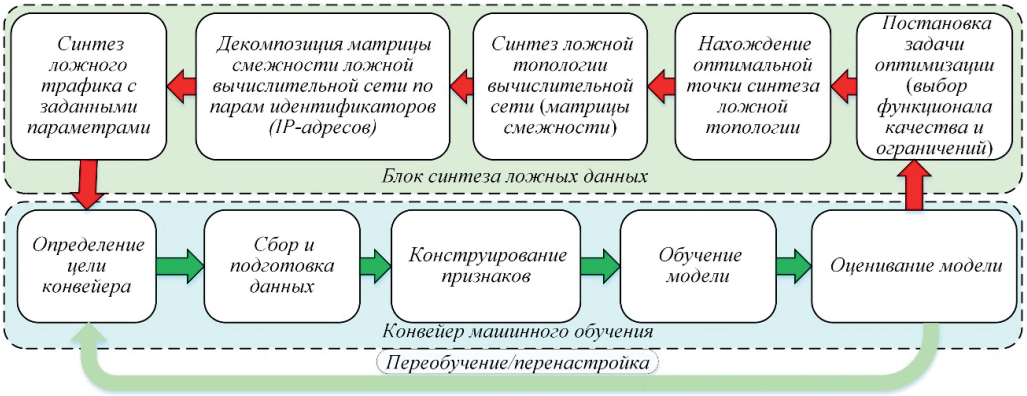

Рис. 2. Схема конвейера машинного обучения и его использование для маскирования топологических свойств вычислительных сетей

Вычислительная сложность. Особую популярность при решении широкого спектра задач получили большие языковые модели, такие как *предобученные генеративные трансформеры* (*Generative Pre-trained Transformer, GPT*). Основной трудностью их использования для маскирования топологических свойств вычислительных сетей является высокая пространственная и временная сложность моделей (количество свободных параметров от $117 \cdot 10^9$ до $500 \cdot 10^9$) [14]. Конвейеры машинного обучения на основе генеративных моделей должны разворачиваться на аппаратных платформах (устройствах, реализующих функцию генератора трафика) с относительно ограниченными возможностями для последующей оптимизации результатов синтеза с использованием численных методов. Непрерывность распределения топологических характеристик в скрытом пространстве обученных моделей ускоряет сходимость и снижает сложность алгоритмов численной оптимизации, определяющих оптимальную точку (или точки) реконструкции топологии вычислительной сети в скрытом пространстве генеративной модели.

Исходя из вышеописанных критериев, в работе используются архитектуры класса «кодировщик-декодировщик», а именно *вариационный сверточный автокодировщик* (*Variotional Autoencoder, VAE*), *вариационный графовый автокодировщик* (*Graph Variotional Autoencoder, GVAE*). Выбранные архитектуры используют разные подходы к снижению размерности и обобщению закономерностей наблюдений: в первом случае, аналогично обработке и генерации изображений, используются сверточные слои; во втором случае, ключевым является слой, который вычисляет нормированный Лапласиан от наблюдений A_i и определяет матрицу из *наименьших собственных векторов* нормированного Лапласиана заданного размера (меньшего, чем исходный размер матриц A_i), то есть основан на матричном разложении матрицы смежности.

Основная идея работы состоит в решении задачи оптимизации в пространстве низкой размерности (скрытом пространстве) генеративных моделей с использованием методов численной оптимизации вещественной скалярной либо векторной целевой функции, при этом выбор оптимальной точки реконструкции топологии осуществляется исходя из критериев, оценивающих *сходство* реальной и ложной топологии, а также смещение ложных критических узлов по отношению к реальным критическим узлам (оценка *защищенности* реальных критических узлов при синтезе ложной топологии).

Реализация маскирования топологических характеристик вычислительных сетей осуществляется на основе *конвейера машинного обучения* с генеративной моделью искусственного интеллекта и *блока синтеза ложных данных*, включающего в себя процедуры постановки задачи оптимизации ложной структуры по выбранным критериям качества, поиска оптимальной точки в скрытом пространстве генеративной модели, обученной на множестве топологических инвариантов вычислительных сетей с заданными свойствами. Далее производится генерация ложной топологии и выделение из полученной матрицы смежности сетевых идентификаторов, в соответствии с которыми осуществляется отправка пакетов ложного сетевого трафика (рис. 2) с заданных интерфейсов устройств, предназначенных для генерации ложного трафика.

Ключевое значение при создании конвейера машинного обучения имеет алгоритм формирования обучающего набора данных, включающий в себя процедуры сбора, подготовки данных и конструирования признаков. Каждое наблюдение, входящее в массив обучающих данных $\mathbf{A}_{train}$, является матрицей смежности реальной вычислительной сети A_{real} размерности N, а также некоторых преобразований $A_i = f(A_{real})$ от матрицы смежности реальной вычислительной сети. Во избежание переобучения

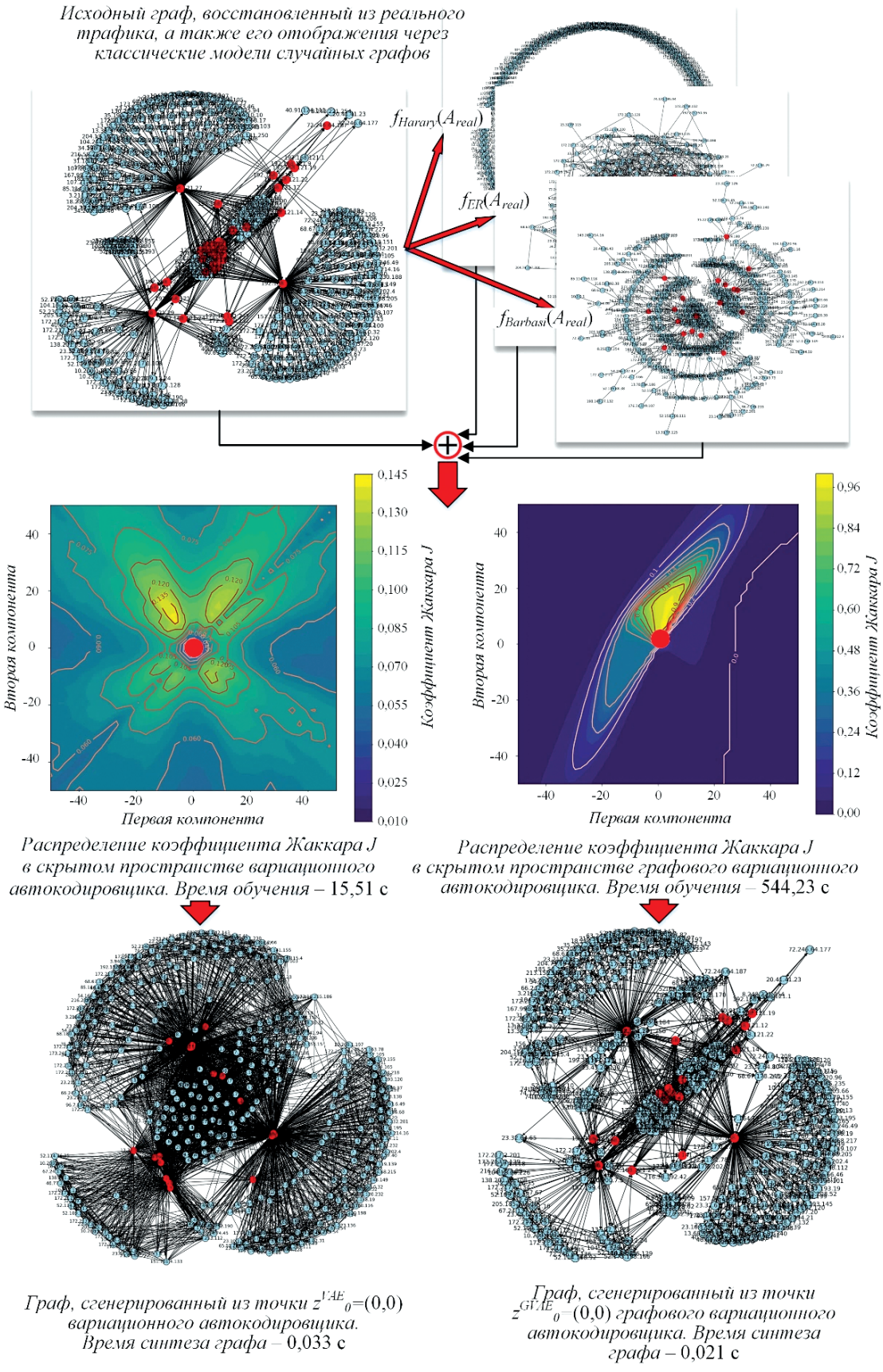

Рис. 3. Схема формирования обучающего набора данных Atrain и реконструкции топологии вычислительной сети с использованием моделей вариационных автокодировщиков

генеративных моделей и для обеспечения достаточного разнообразия ложных топологий вычислительных сетей, при формировании обучающего набора данных используются матрицы смежности A_i, полученные из моделей случайных графов (*Эрдеша-Реньи, Ватца-Строгаца, Барбаши и Харари*).

Формирование обучающего датасета $\mathbf{A}_{train}$ производится за счет конкатенации и перемешивания матриц смежности A_i, полученных из различных моделей случайных графов, а также матриц смежности A_{real} реальной вычислительной сети в равных соотношениях (рис. 3).

При этом точечные оценки параметров θ^{RG}_i указанных моделей случайных графов определяются исходя из максимизации среднего коэффициента Жаккара между множеством ребер реальной E_{real} и множествами ребер ложных вычислительных сетей E_i, полученных с помощью параметризованных оценками θ^{RG}_i моделей случайных графов.

При обучении вариационного автокодировщика решается задача оптимальной реконструкции входных данных кодировщика из выходных данных декодировщика с учетом близости распределения наблюдений в скрытом пространстве кодера к многомерному нормальному при фиксированных оптимальных значениях гиперпараметров θ^*_{gip}. Перед развертыванием конвейера машинного обучения осуществляется гиперпараметрическая оптимизация модели. Для архитектур вариационных автокодировщиков часто используют нижнюю вариационную границу (*Evidence Lower Bound*) в качестве ненормированной взвешенной линейной свертки как функционала качества обучения и гиперпараметрической оптимизации модели (выражение 1) [15]:

$$L(\theta_{gip},\theta,\varphi,\mathbf{A}_{train}) = \\ = k_1 D_{KL}(N(\mu,\sigma\|N(0,1)) + k_2 MSE \longrightarrow \min_{Q}, \quad (1)$$

$$MSE(\theta_{gip},\theta,\varphi,\mathbf{A}_{train}) = \frac{1}{n}\sum_{i=1}^{n}(A_i - \hat{A}_i(\theta_{gip},\theta,\varphi,\mathbf{A}_{train}))^2, \quad (2)$$

$$\mu = f(\theta_{gip},\theta,\varphi,\mathbf{A}_{train}),\ \sigma = f(\theta_{gip},\theta,\varphi,\mathbf{A}_{train}), \quad (3)$$

$$D_{KL}(N(\mu,\sigma\|N(0,\mathbf{I})) = \\ = \frac{1}{2N_{Lat}}\sum_{i=1}^{N_{Lat}}(1 + \log(\sigma_i^2) - \mu_i^2 - \sigma_i^2), \quad (4)$$

$$\theta^*_{gip},\theta^*,\varphi^* = \arg\min_{Q} L(\theta_{gip},\theta,\varphi,\mathbf{A}_{train}), \quad (5)$$

где, k_1, k_2 – коэффициенты значимости целевых функций; θ_{gip} – множество гиперпараметров генеративной модели (тип и количество нейронов, количество слоев кодера/декодера, активационная функция, количество эпох обучения, размер обучающей выборки, размер пакетов); θ, φ – свободные параметры кодировщика и декодировщика (веса и смещения нейронов); $\mathbf{A}_{train}$ – тренировочный массив данных с матрицами смежности исходных топологий вычислительных сетей; D_{KL} – дивергенция Кульбака-Лейблера между распределением представлений графов в скрытом пространстве $Z \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma})$ и многомерным стандартным нормальным распределением $N(\mathbf{0}, \mathbf{I})$; MSE – среднее квадратическое отклонение между исходными матрицами смежности A_i из тензора A_{real} и реконструируемыми матрицами смежности $\hat{A}i$; Q – допустимое множество значений целевых функций и аргументов; N_{Lat} – размерность скрытого пространства Z генеративной модели; n – количество наблюдений (матриц смежности) в пакете с обучающими данными.

Архитектуры моделей, оптимальные гиперпараметры θ^*_{gip} сверточного и графового вариационных автокодировщиков, используемых для синтеза ложной топологии вычислительной сети, состоящей из 392 вершин, представлены на рис. 4.

В соответствии с замыслом работы показателем защищенности реальной вычислительной сети при генерации ложной топологии вычислительной сети является *аппроксимация дистанции между множеством реальных и ложных критических узлов*. В качестве данной характеристики используется среднее кратчайшее расстояние D между критическими ложными и реальными узлами, а также коэффициент Жаккара J_{crit} между множеством критических ложных и исходных узлов. То есть, защищенность реальной вычислительной сети повышается если в среднем реальные критические узлы расположены дальше от ложных критических узлов, при этом отсутствуют пересечения между множеством ложных и реальных критических узлов. При рассмотрении только топологических характеристик, расстояние D рассчитывается в *хопах* (*скачках*), то есть в среднем минимальном количестве промежуточных узлов между множествами критических узлов.

Правдоподобие или сходство генерируемой ложной топологии вычислительной сети оценивается с помощью функционала сходства ложной и реальной топологии, в качестве которого используется коэффициент Жакарра J_{edge} между множеством ребер ложной и реальной вычислительной сети.

В соответствии с заданными критериями качества ложных топологий выбор оптимальных точек из скрытого пространства вариационного автокодировщика производится на основе решения следующей задачи многокритериальной оптимизации (выражение 6). Решение данной задачи в общем случае имеет множество Парето оптимальных решений. Использование *метаэвристических алгоритмов* получило широкое распространение при решении задач многокритериальной оптимизации, в частности, популяционные, эволюционные и смешанные

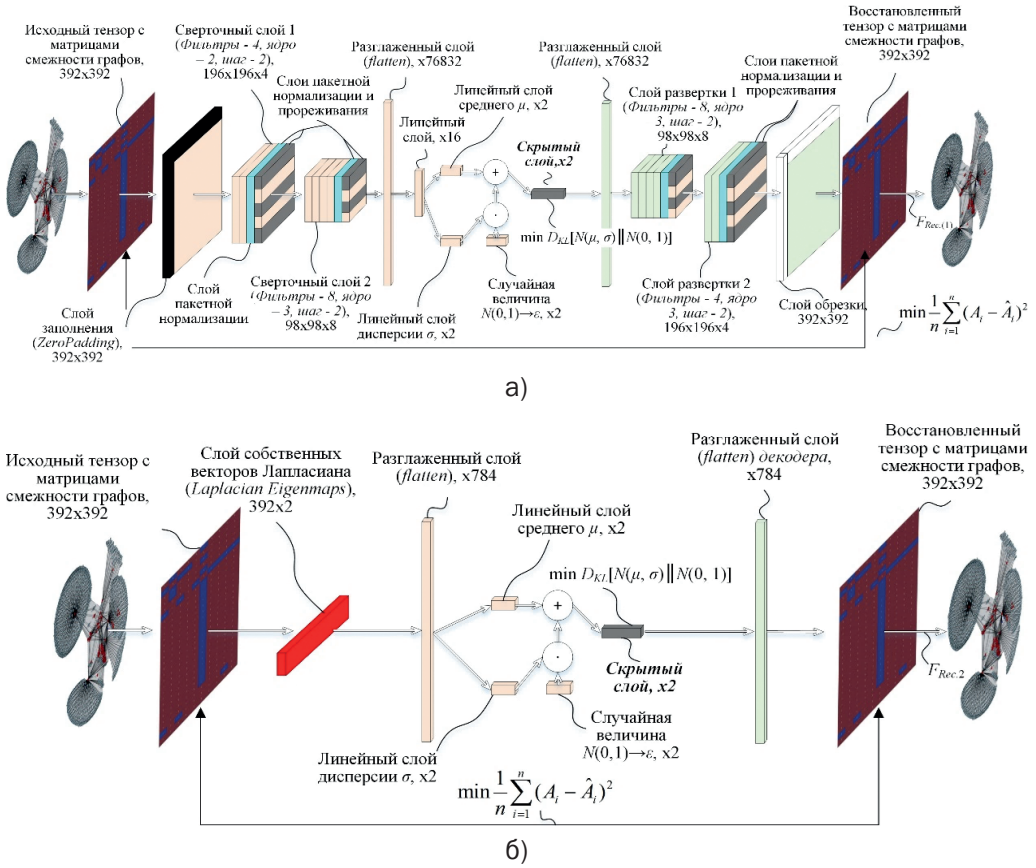

Рис. 4. Архитектуры вариационных автокодировщиков с оптимальными гиперпараметрами, предназначенные для синтеза матрицы смежности ложной вычислительной сети с заданными характеристиками: а) сверточный вариационный автокодировщик, б) графовый вариационный автокодировщик; где, F_{Rec1}, F_{Rec2} – функции округления выходных тензоров

алгоритмы [16–18]. Выбор конкретного алгоритма осуществляется на основе априорной информации о свойствах целевых функций, ограничений и требований к вычислительной сложности.

$$\begin{cases} J_{edge}(\theta_{hid}, A_{real}) \longrightarrow \max\limits_{Q_2}, \\ D(\theta_{hid}, A_{real}) \longrightarrow \max\limits_{Q_2}, \\ J_{crit}(\theta_{hid}, A_{real}) \longrightarrow \min\limits_{Q_2}. \end{cases} \quad (6)$$

где, $\theta_{hid} = (\theta^{hid}_1, \ldots, \theta^{hid}_m)$ – координаты скрытого пространства вариационного автокодировщика в пространстве R^{Nlat}, используемые для синтеза оптимальный топологии ложной вычислительной сети A; Q_2 – допустимое множество значений целевых функций и аргументов.

Рассмотрим упрощенный случай, в котором скрытое пространство генеративных моделей имеет всего два измерения, а решение задачи многокритериальной оптимизации осуществляется посредством сведения к скалярной функции. Поиск точки синтеза (x^*_1, x^*_2) ложной структуры в скрытом пространстве R^2 генеративной модели осуществлялся посредством минимизации функционала качества f, представленного в форме ненормированной взвешенной линейной свертки критериев (выражения 7–9):

$$f = \alpha_1 \cdot \bar{J}_{edge}(x_1, x_2) + \\ + \alpha_2 \cdot (D(x_1, x_2) + \varepsilon)^{-1} + \alpha_3 \cdot J_{crit}(x_1, x_2), \quad (7)$$

$$f \longrightarrow \min\limits_{Q_3}, \ (x^*_1, x^*_2) = \arg\min\limits_{Q_2} f(x_1, x_2), \quad (8)$$

$$Q_3 = \begin{cases} \alpha_1 \in [0,1], \ \alpha_2 \in [0,1], \ \alpha_3 \in [0,1], \\ \bar{J}_{edge}(x_1, x_2) \in [0,1], \ \varepsilon = 1{,}0, \\ J_{crit}(x_1, x_2) \in [0,1], \ D(x_1, x_2) \geq 0, \\ x_1, x_2 \in [-50, 50]. \end{cases} \quad (9)$$

где, α_1, α_2, α_3 – коэффициенты значимости целевых функций; x_1, x_2 – координаты точки в скрытом пространстве R^2; ε – поправочный коэффициент, предотвращающий деление на 0.

В общем случае характер распределения значений целевых функций L (выражения 1–5) и f (выражения 7–9), количество локальных экстремумов, обусловленность моделей, неизвестны, несмотря на то, что распределения точек скрытого пространств в генеративных моделях близки к стандартному

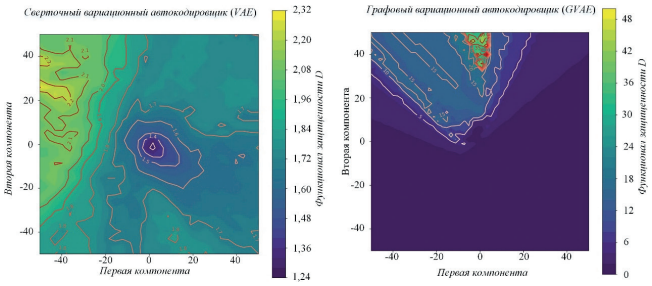

Рис. 5. Распределение среднего кратчайшего расстояния между реальными и ложными критическими узлами D в скрытых пространствах обученных генеративных моделей

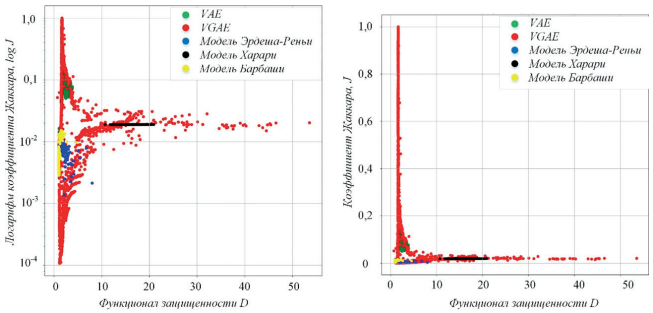

Рис. 6. Сравнение допустимого множества в критериальном пространстве показателей качества классических моделей случайных графов и вариационных автокодировщиков

нормальному. В работе решение указанных задач оптимизации осуществлялось с использованием алгоритма байесовской оптимизации, имеющего приемлемую сложность и сходимость для многопараметрических задач.

Распределения показателя защищенности D топологий ложных вычислительных сетей, синтезируемых из соответствующих точек двумерных скрытых пространств генеративных моделей, характеризуются значительными различиями как в характере распределений, так и в диапазонах значений. Так для сверточного вариационного автокодировщика среднее кратчайшее расстояние D находится в диапазоне от 1,24 до 2,32 хопов, при этом минимальные значения сгруппированы в окрестности центральной точки с координатами (0, 0), а для архитектуры графового вариационного автокодировщика наибольшие значения показателя D распределены дальше от центральной точки, при этом имеет место более широкий диапазон значений от 0 до 48 хопов (рис. 5).

Для сравнительной характеристики качества различных генеративных моделей на рис. 6 изображены допустимые множества соотношений первых двух критериев качества ложных топологий в двумерном критериальном пространстве для моделей Эрдеша-Реньи, Барбаши, Харари, сверточного и графового вариационных автокодировщиков. Стоит отметить, что при использовании моделей безразмерных графов топологические характеристики синтезируемых графов распределены в относительно узких диапазонах, однако, вычислительная сложность данных моделей является очень низкой, что позволяет их использовать для синтеза топологий с количеством вершин более 10^3.

С другой стороны, модели искусственного интеллекта показывают значительно более широкий диапазон характеристик качества структур, причем для архитектуры графового вариационного автокодировщика имеет место наибольший диапазон возможных топологических характеристик синтезируемых структур. Тем не менее с связи со сложностью матричных разложений данная архитектура является наиболее требовательной к вычислительным ресурсам.

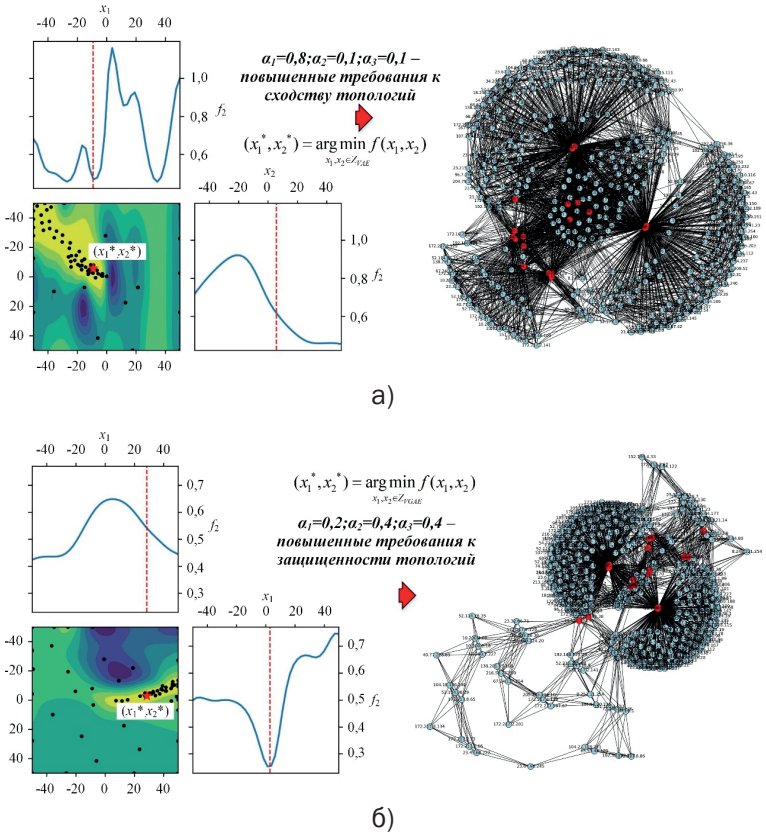

Рис. 7. Результаты решения задачи выбора оптимальной точки реконструкции ложной топологии вычислительной сети с использованием модели графового вариационного автокодировщика: а) при значениях коэффициентов значимости α_i, повышающих предпочтение к сходству ложной топологии; б) при значениях коэффициентов значимости α_i, повышающих предпочтение к защищенности критических узлов

Практическое применение. Генерация ложного сетевого трафика может быть реализована на различных устройствах, поддерживающих возможности отправки сообщений с заданными сетевыми параметрами (маршрутизаторы, межсетевые экраны, компьютеры). Ресурсы устройств, выполняющих вычислительную задачу по синтезу ложной топологии, влияют на выбор архитектуры, оптимальных значений гиперпараметров и свободных параметров генеративных моделей. Ложные сетевые информационные объекты (приманки, ловушки) в качестве источников ложного трафика могут быть развернуты на базе одного (с помощью управляемых виртуальных сетевых интерфейсов) или нескольких серверов. При ограниченных вычислительных ресурсах (генераторы на основе микрокомпьютеров и объемом оперативной памяти до 4 Гбайт) для синтеза ложной структуры целесообразно использовать модели со значительно меньшим количеством параметров (эвристические алгоритмы, модели случайных графов и ансамбли из простейших моделей случайных графов). Также модели с низкой пространственной сложностью целесообразно использовать в случаях, когда сетевой трафик содержит тысячи узлов и десятки тысяч ребер, так как обучающие тензоры $\mathbf{A}_{train}$ будут требовать значительных объемов оперативной памяти, в связи с тем, что асимптотические оценки сложности рассмотренных генеративных моделей, при фиксированных параметрах сверточных слоев и размера матрицы собственных векторов Лапласиана, составляют $O(N^2)$.

Вывод

Для маскирования топологических характеристик вычислительных сетей относительно большой размерности (при $10^2 < N < 10^3$) целесообразно использовать модели и методы генеративного искусственного интеллекта на вычислительных устройствах, способных обеспечить работоспособность конвейеров машинного обучения:

- архитектуры графового и сверточного вариационных автокодировщиков позволяют обрабатывать немаркированные данные из сетевого трафика, обладают относительно невысокой пространственной и временной сложностью при рассмотренных ограничениях на размерность составной сети, имеют возможность синтезировать ложные топологии вычислительных сетей в широких диапазонах показателей качества;
- модель выборки, задающая характеристики ложной топологии, может включать коэффициент Жаккара между множеством ребер реальной и ложной топологии в качестве показателя сходства, коэффициент Жаккара между множеством ложных и реальных критических узлов, а также среднее кратчайшее расстояние между реальными и ложными критическими узлами в качестве показателей защищенности;
- для поиска оптимальных точек синтеза ложной топологии в скрытом пространстве генеративных моделей целесообразно использовать метаэвристические алгоритмы скалярной и многокритериальной оптимизации, в частности, алгоритм байесовской оптимизации;
- модели случайных графов целесообразно использовать на этапе формирования набора данных для обучения генеративных моделей либо в качестве генераторов ложных топологий с количеством вершин $N > 10^3$.

Литература

1. *Горбачев А. А., Максимов Р. В. Проблема маскирования и применения технологий машинного обучения в киберпространстве // Вопросы кибербезопасности. 2023. № 5 (57). С. 37–49. DOI: 10.21681/4311-3456-2023-5-37-49.*
2. *Москвин А. А., Максимов Р. В., Горбачев А. А. Модель, оптимизация и оценка эффективности применения многоадресных сетевых соединений в условиях сетевой разведки // Вопросы кибербезопасности. 2023. № 3 (55). С. 13-22. DOI: 10.21681/2311-3456-2023-3-13-22.*
3. *Maximov R. V., Sokolovsky S. P., Telenga A. P. Methodology for sustaniating the characteristics of false network traffic to simulate information systems // Selected Papers of the XI International Scientific and Technical Conference on Secure Information Technologies (BIT-2021). 2021. p. 115–124.*
4. *Maximov R. V., Sokolovsky S. P., Telenga A. P. Honeypots network traffic parameters modelling // Selected Papers of the XI International Scientific and Technical Conference on Secure Information Technologies (BIT-2021). 2021. p. 229–239.*
5. *Кузьмин В. Н., Шуваев Ф. Л., Розганов М. В. Сравнительный анализ моделей случайных графов // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. №. 58. С. 23–34.*
6. *Лыгин В. С., Сирота А. А., Головинский П. А. Регуляризация процесса обучения графовых нейронных сетей методом распространение меток // Вестник ВГУ. Серия: Системный анализ и информационные технологии. 2024. №. 3. С. 92–101. DOI: 10.17308/sait/1995-5499/2024/3/92-101.*
7. *Schweinberger M., Krivitsky P. N., Butts C. T., Stewart J. R. Exponential-Family Models of Random Graphs: Inference in Finite, Super and Infinite Population Scenarios. Statistical Science. 2020. Vol. 35. No. 4. pp. 627–662. DOI: 10.1214/19-STS743.*
8. *Fanourakis N., Efthymiou V., Kotzinos D., Christophides V. Knowledge graph embedding methods for entity alignment: experimental review. Data Mining and Knowledge Discovery. 2023. Vol. 37. pp. 2070–2137. DOI: 10.1007/s10618-023-00941-9.*
9. *Said A., Shabbir M., Hassan S., Hassan Z. R., Ahmed A., Koutsoukos X. On augmenting topological graph representations for attributed graphs. Applied Soft Computing. 2023. Vol. 136. 110104. DOI: 10.1016/j.asoc.2023.110104.*
10. *Van Der Hofstad R. Random graphs and complex networks. Cambridge university press. 2024. Volume 2. 492 p.*

11. *Xu M. Understanding Graph Embedding Methods and Their Applications. Society for Industrial and Applied Mathematics. 2021. Vol. 63. No 4. pp. 825–853. DOI: 10.1137/20M1386062.*
12. *Li J., Fu X., Sun Q., Ji C., Tan J., Wu J., Peng H. Curvature graph generative adversarial networks. In Proceedings of the ACM web conference 2022. 2022. pp. 1528–1537. DOI: 10.1145/3485447.3512199.*
13. *Naveed H. et al. A comprehensive overview of large language models // ArXiv. 2023. pp. 1–35.*
14. *Коробцов В.И., Овсянников И.В., Сачков Д.И. Автоматическая генерация надежного программного кода с помощью генеративных предобученных трансформеров (GPT) // «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. 2024. №1. С.52–59.*
15. *Mrabah N., Bouguessa M., Ksantini R. Beyond The Evidence Lower Bound: Dual Variational Graph Auto-Encoders For Node Clustering. In Proceedings of the 2023 SIAM International Conference on Data Mining (SDM). 2023. pp. 100–108.*
16. *Sharma S., Kumar V. A comprehensive review on multi-objective optimization techniques: Past, present and future. Archives of Computational Methods in Engineering. 2022. Vol. 29(7). pp. 5605–5633. DOI: 10.1007/s11831-022-09778-9.*
17. *Asfar B., Miettinen K., Ruiz F. Assessing the performance of interactive multiobjective optimization methods: A survey. ACM Computing Surveys (CSUR). 2021. Vol. 54(4). pp. 1–27. DOI: 10.1145/3448301.*
18. *Liu S., Lin Q., Wong K. C., Li Q., Tan K. C. Evolutionary large-scale multiobjective optimization: Benchmarks and algorithms. IEEE Transactions on Evolutionary Computation. 2021. Vol. 27(3). pp. 401–415. DOI: 10.1109/TEVC.2021.3099487.*

MASKING THE TOPOLOGICAL PROPERTIES OF COMPUTER NETWORKS IN THE CONDITIONS OF NETWORK RECONNAISSANCE. Part 2

Gorbachev A. A.[7]

Keywords: *false information objects, variational autoencoder, machine learning pipeline, artificial intelligence, optimization, metaheuristic algorithms, random graphs.*

The purpose of the study: *to develop a model system including classical random graph models and generative artificial intelligence models designed to solve the problem of masking the topological properties of computer networks when generating false network traffic and using false network information objects, allowing on the one hand to ensure a given degree of similarity of the topological properties of real computer networks with false ones, and on the other hand to maximize an indicator of the security of critical nodes of real computer networks.*

Methods used: *Erdos-Renyi random graph, Barbashi, Watts-Strogatz, Harari, Bayesian optimization algorithm, convolutional variational autoencoder model, graph variational autoencoder model, weighted additive linear convolution.*

The result of the study: *the presented system of models makes it possible to increase the effectiveness of protecting a computer network by forming a stable false idea in an attacker about the topological properties of a computer network, taking into account the increased security of critical nodes by shifting the position of false critical nodes relative to real ones, while ensuring a given degree of similarity of the false topology of a computer network in relation to the real topology. The model system includes a machine learning pipeline based on random graph models of Erdos-Renyi, Barbashi, Watts-Strogatz, Harari, used to form a training dataset, a graph variational autoencoder model, a hidden space sampling model containing quality indicators of the generated false structure, an evolutionary scalar optimization algorithm that searches for the optimal synthesis point a false structure in the hidden space of a variational auto-encoder, as well as a false traffic generator, which implements sending packets with the specified network identifiers. The developed pipeline has limitations in the dimension of the synthesized false topology due to the computational complexity of the generative model learning process and the search for the optimal synthesis point.*

Scientific novelty: *it consists in the application of a Bayesian optimization algorithm to select the optimal point for the synthesis of a false topology from the hidden space of a trained graph variational autoencoder, in using the objective function represented by a linear weighted convolution from the Jacquard coefficient between the set of edges of the false and real topology of the computer network, indicators of the security of the computer network: the average shortest distance between real and false critical nodes, the Jacquard coefficient between the set of false and real critical nodes of a computer network. In the application of random graph models to form a training dataset.*

References

1. *Gorbachev A.A., Maksimov R.V. Problema maskirovaniya i primeneniya texnologij mashinnogo obucheniya v kiberprostranstve // Voprosy` kiberbezopasnosti. 2023. № 5 (57). S. 37–49. DOI:10.21681/4311-3456-2023-5-37-49.*

7 Gorbachev Alexander, candidate of technical sciences. Krasnodar Higher Military Orders of Zhukov and the October Revolution Red Banner School named after General of the Army S.M. Shtemenko, Krasnodar, Russia. E-mail: infosec23.00@mail.ru

2. *Moskvin A.A., Maksimov R.V., Gorbachev A.A. Model', optimizaciya i ocenka e'ffektivnosti primeneniya mnogoadresny'x setevy'x soedinenij v usloviyax setevoj razvedki // Voprosy` kiberbezopasnosti. 2023. № 3 (55). S. 13-22. DOI: 10.21681/2311-3456-2023-3-13-22.*
3. *Maximov R.V., Sokolovsky S.P., Telenga A.P. Methodology for sustaniating the characteristics of false network traffic to simulate information systems // Selected Papers of the XI International Scientific and Technical Conference on Secure Information Technologies (BIT-2021). 2021. p. 115–124.*
4. *Maximov R.V., Sokolovsky S.P., Telenga A.P. Honeypots network traffic parameters modelling // Selected Papers of the XI International Scientific and Technical Conference on Secure Information Technologies (BIT-2021). 2021. p. 229–239.*
5. *Kuz'min V.N., Shuvaev F.L., Rozganov M.V. Sravnitel'ny'j analiz modelej sluchajny'x grafov // Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vy`chislitel`naya texnika i informatika. 2022. №. 58. S. 23–34.*
6. *Ly'gin V.S., Sirota A.A., Golovinskij P.A. Regulyarizaciya processa obucheniya grafovy`x nejronny`x setej metodom rasprostranenie metok // Vestnik VGU. Seriya: Sistemny`j analiz i informacionny`e texnologii. 2024. №. 3. S. 92–101. DOI: 10.17308/sait/1995-5499/2024/3/92-101.*
7. *Schweinberger M., Krivitsky P.N., Butts C.T., Stewart J.R. Exponential-Family Models of Random Graphs: Inference in Finite, Super and Infinite Population Scenarios. Statistical Science. 2020. Vol. 35. No. 4. pp. 627–662. DOI: 10.1214/19-STS743.*
8. *Fanourakis N., Efthymiou V., Kotzinos D., Christophides V. Knowledge graph embedding methods for entity alignment: experimental review. Data Mining and Knowledge Discovery. 2023. Vol. 37. pp. 2070–2137. DOI: 10.1007/s10618-023-00941-9.*
9. *Said A., Shabbir M., Hassan S., Hassan Z.R., Ahmed A., Koutsoukos X. On augmenting topological graph representations for attributed graphs. Applied Soft Computing. 2023. Vol. 136. 110104. DOI: 10.1016/j.asoc.2023.110104.*
10. *Van Der Hofstad R. Random graphs and complex networks. Cambridge university press. 2024. Volume 2. 492 p. DOI: 10.1137/20M1386062.*
11. *Xu M. Understanding Graph Embedding Methods and Their Applications. Society for Industrial and Applied Mathematics. 2021. Vol. 63. No 4. pp. 825–853. DOI: 10.1145/3485447.3512199.*
12. *Li J., Fu X., Sun Q., Ji C., Tan J., Wu J., Peng H. Curvature graph generative adversarial networks. In Proceedings of the ACM web conference 2022. 2022. pp. 1528–1537.*
13. *Naveed H. et al. A comprehensive overview of large language models // ArXiv. 2023. pp. 1–35.*
14. *Korobczov V.I., Ovsyannikov I.V., Sachkov D.I. Avtomaticheskaya generaciya nadezhnogo programmnogo koda s pomoshh'yu generativny'x predobuchenny'x transformerov (GPT) // «Informacionny'e texnologii i matematicheskoe modelirovanie v upravlenii slozhny'mi sistemami»: e'lektron. nauch. zhurn. 2024. №1. S. 52–59.*
15. *Mrabah N., Bouguessa M., Ksantini R. Beyond The Evidence Lower Bound: Dual Variational Graph Auto-Encoders For Node Clustering. In Proceedings of the 2023 SIAM International Conference on Data Mining (SDM). 2023. pp. 100–108.*
16. *Sharma S., Kumar V. A comprehensive review on multi-objective optimization techniques: Past, present and future. Archives of Computational Methods in Engineering. 2022. Vol. 29(7). pp. 5605–5633. DOI: 10.1007/s11831-022-09778-9.*
17. *Asfar B., Miettinen K., Ruiz F. Assessing the performance of interactive multiobjective optimization methods: A survey. ACM Computing Surveys (CSUR). 2021. Vol. 54(4). pp. 1–27. DOI: 10.1145/3448301*
18. *Liu S., Lin Q., Wong K.C., Li Q., Tan K.C. Evolutionary large-scale multiobjective optimization: Benchmarks and algorithms. IEEE Transactions on Evolutionary Computation. 2021. Vol. 27(3). pp. 401–415. DOI: 10.1109/TEVC.2021.3099487.*