

А.А. АНДРЕЕВ, А.И. ШАБАЕВ  
**МОДЕЛИ И МЕТОДЫ ВЫЯВЛЕНИЯ СТРУКТУРЫ  
ЛОКАЛЬНОЙ ВЫЧИСЛИТЕЛЬНОЙ СЕТИ ПРИ НЕПОЛНЫХ  
ДАННЫХ**

*Андреев А.А., Шабаетв А.И. Модели и методы выявления структуры локальной вычислительной сети при неполных данных.*

**Аннотация.** Описание логической и физической структур сети необходимо для решения многих задач сетевого управления. Автоматизация такого описания осложнена неполнотой и некорректностью исходных данных о структуре сети, получаемых из стандартных источников. Это приводит к необходимости выявления скрытых и неявно заданных элементов структуры сети. Исследуются свойства неполных исходных данных о связях сетевых устройств на канальном уровне. Для этого вводятся графовые модели структуры канального уровня в присутствии виртуальных локальных сетей, формализуется задача дополнения исходных данных. Приводятся методы обобщенной обработки разнородных исходных данных о канальном уровне. Описываются модели и методы выведения части отсутствующих данных, а также условие, при котором исходные данные могут позволить построить единственное правильное описание структуры сети. Рассматриваются методы определения в исходных данных некорректных элементов и ситуаций, при которых возможно несколько решений задачи выявления структуры сети. Приводятся алгоритмы выявления структуры сети при неполных данных, когда их можно и нельзя восполнить до необходимого объема, методы разрешения неоднозначности во входных данных и методы исправления некорректных данных. Проведенные вычислительные испытания методов на сгенерированных данных и в реальных сетях показывают их применимость и эффективность в сетях различных структур и со смешанным набором производителей оборудования. Также показывается преимущество перед разработанными ранее аналогичными методами: способность дополнить до 99% данных о связях на канальном уровне за полиномиальное время, возможность получения точного решения при неоднозначности исходных данных.

**Ключевые слова:** компьютерные сети, моделирование, автоматизация, структура ЛВС, канальный уровень

**1. Введение.** Развитие рынка информационных технологий обуславливает рост масштабов и сложности коммуникационных инфраструктур (далее сети) современных поставщиков сетевых услуг (ПСУ). Самым многочисленным классом ПСУ являются предприятия малой и средней величины, решающие информационно-технологические задачи на базе собственной локальной вычислительной сети (ЛВС). Систематический подход к управлению требует полного и детального описания структуры ЛВС, включая все три базовых уровня (физический, канальный и сетевой).

Подробное описание структуры ЛВС, включающее сведения об управляемых элементах (сетевые устройства и порты, конечные точки протоколов передачи данных) и связях иерархии и передачи данных между ними, используется сетевыми администраторами для моделирования, про-

ектирования и документирования сети, обеспечения отказоустойчивости и безопасности, анализа производительности, построения оптимальной структуры сети с точки зрения надежности и количества соединений [1–4].

Из-за масштабов современных ЛВС построение и сопровождение такого описания вручную трудозатратно, в связи с чем встает задача автоматизации описания структуры функционирующей сети по доступным динамическим данным [5, 6]. Сложность решения этой задачи заключается в неполноте исходных данных [5–8], которая возникает из-за устаревания данных (например, в кэше Address Resolution Protocol ARP и таблицах коммутации), разнородности сетевого оборудования, отсутствия части данных по техническим причинам и мерам безопасности. Эти и другие факторы обуславливают необходимость выявления скрытых и неявно заданных элементов структуры сети.

Известно, что неполнота данных может приводить к ошибкам и неточностям в автоматически построенных описаниях структуры сети [5, 7]. Однако можно вывести часть отсутствующих данных и сделать процесс исследования сети более точным [5, 7, 9, 10].

Для решения задачи автоматизации выявления структуры сети в серии предыдущих работ [9–12] представляют модели и методы автоматизированного выявления структуры ЛВС на физическом, канальном и сетевом уровнях при неполных данных и присутствии прозрачных устройств. В частности, были описаны методы восполнения части данных о связях на канальном уровне, исследованы свойства предложенных решений [9, 10]. Ограничением работоспособности предложенных методов является необходимость корректности исходных данных и наличия в них сведений о достижимости между собой на канальном уровне всех соединенных физически устройств [9]. Цель данной статьи – исследование методов дополнения неполных исходных данных и построение новых методов, способных работать при обозначенных ограничениях.

Статья организована следующим образом. В параграфе 2 представлен обзор ранее выполненных исследований по теме работы, выявлены некоторые недостатки. В параграфе 3 представлена графовая модель структуры сети с учетом VLAN (Virtual Local Area Network), а также модель данных о связях на канальном уровне сети. Представлены основные понятия, используемые при выявлении структуры сети при неполных исходных данных. В параграфе 4 представлены алгоритмы обработки неполных исходных данных в рамках процесса выявления структуры сети. В параграфе 5 описаны результаты испытаний предложенных методов. Параграф 6 завершает статью.

**2. Обзор литературы.** Проблема неполноты исходных данных тщательно исследуется в задаче автоматизации выявления структуры ЛВС. Эта проблема больше всего касается данных о связях устройств на канальном уровне, которые лежат в основе большинства методов автоматизированного описания структуры сети. В частности, таблицы коммутации (Address Forwarding Tables, AFT), которые широко используются для исследования сети, зависят от активности сетевых устройств и содержат информацию не о прямых (физических) соединениях, а о достижимости, то есть о возможности обмена между интерфейсами кадров канального уровня.

В [13] указывается на необходимость глубокого исследования свойств таблиц коммутации в связи с их возможной неполнотой. Описывается модель и набор правил для обработки неполных данных и выявления структуры сети. На практике показана способность метода к построению точного описания структуры сети при наличии 15% данных о достижимости. Однако описанный метод применим только к сетям с деревообразной структурой и без VLAN. Также при выявлении структуры сети возможно возникновение конфликтных ситуаций, для разрешения которых предложены полуслучайные методы, что может привести к большому количеству ошибок.

Исследователи в [7] описали некоторые общие свойства неполных данных о достижимости и предложили алгоритм их обработки для использования при выявлении структуры сети. Доказана достаточность полученных дополненных данных для исследования сети. Однако данный метод не может работать в сетях с VLAN и может дать некорректный результат при присутствии в сети прозрачных устройств (например, хабов без MAC-адресов).

В [5] показано, что выявление структуры сети при неполных исходных данных, а также определение, могут ли исходные данные привести к однозначному решению, являются NP-полными задачами от числа сетевых узлов. В статье предложены модели структуры сети и неполных входных данных, а также набор правил и алгоритмов для выявления структуры сети в различных ситуациях. Вычислительные испытания одного из предложенных алгоритмов показали, что он способен строить точное описание структуры сети при полноте данных более 50%. Однако алгоритмы не поддерживают сети с VLAN и часто могут дать неоднозначное решение.

В [6, 8] исследователями предприняты попытки адаптировать результаты из [5] к сетям с VLAN. В [6] авторами показано, что использование нескольких источников данных, а не только таблиц коммутации, может повысить точность выявления структуры сети и полноту исходных

данных. Для этого авторы в дополнение используют ARP-кэш как источник данных о достижимости на канальном уровне. В [8] показано, что можно ослабить требования методов из [6] на входные данные: достаточно определенной степени полноты данных о связях между служебными сетевыми устройствами (исключая хосты). Однако ни одна из этих работ не предоставляет теоретического анализа или подробных вычислительных испытаний своих методов при неполных исходных данных. Утверждается только, что удалось построить описание структуры сети в нескольких частных случаях.

Методы, представленные в [14], способны работать в сетях с VLAN и с неполными данными. Для этого авторы объединяют построение описания канального и сетевого уровней. Однако авторами не приводятся исследования влияния неполноты данных на результат. Кроме того, приведенные методы не работоспособны в присутствии прозрачных устройств. Также авторами упоминается, что не гарантируется корректная работа алгоритма в сетях, в которых используется оборудование нескольких производителей.

Во многих работах [15–20] проблема неполноты данных упоминается, но никак не решается. То есть модели и методы выявления структуры сети, предлагаемые в этих работах, проектируются только для случаев с полными исходными данными, чего нельзя гарантировать в реальных сетях.

В [9, 10] авторами разработаны модели и методы выявления структуры при неполных исходных данных из нескольких источников различных типов вне зависимости от производителей оборудования. Показана теоретическая и практическая применимость и эффективность методов при низкой полноте исходных данных. Показано, какие данные должны присутствовать в исходном наборе для однозначного решения задачи выявления структуры сети. Однако в работе не рассматриваются случаи, когда исходных данных недостаточно для получения однозначного решения или они некорректны.

Новизна представленных в настоящей работе результатов заключается в определении и исследовании новых свойств задачи обработки неполных данных о структуре сети. Разработан новый метод дополнения неполных данных и определено формальное условие, когда этот метод может дать точное решение. Также представлен ряд методов повышения точности приближенного решения при невозможности получения точного.

**3. Модель структуры сети.** В [10, 12] описана графовая модель структуры сети на физическом, канальном и сетевом уровнях с учетом VLAN, VPN и агрегирования каналов, описаны ее свойства. Приведем

краткие сведения об этой модели на канальном уровне, что потребуется для описания разрабатываемых методов.

**3.1. Группированные графы.** Заметим, что структура и свойства графа на трех описываемых уровнях схожая и приведем обобщение некоторых их свойств.

Группированным графом назовем граф  $G = \langle V, E \rangle = \langle O \cup P, E \rangle$ , в котором каждая вершина  $o \in O$  является центром подграфа-звезды (группы) с подмножеством вершин  $P_o \subset P$  и для любых  $o_1, o_2 \in O$  ребро  $(o_1, o_2) \notin E$ . Вершины множества  $O$  назовем базовыми вершинами группированного графа.

Введем несколько классов отношений (ребер) в группированном графе. Отношением ассоциации  $A$  назовем бинарное симметричное антирефлексивное отношение на множестве  $O \cup P$  между базовой вершиной и небазовой вершиной из ее группы:  $(p, o) \in A$  и  $(o, p) \in A$  тогда и только тогда, когда  $p \in P, o \in O$  и  $p \in P_o$ . Отношением соединения назовем бинарное симметричное слабо транзитивное (транзитивность без рефлексивности) антирефлексивное отношение на множестве  $P$  между небазовыми вершинами из разных групп. Отношением перенаправления назовем бинарное симметричное слабо транзитивное антирефлексивное отношение на множестве  $P$  между небазовыми вершинами одной группы.

Многоуровневым группированным графом назовем граф  $G = \langle V, E \rangle = G^{(1)} \cup G^{(2)} \cup \dots \cup G^{(n)} = \langle V^{(1)} \cup \dots \cup V^{(n)}, E^{(1)} \cup \dots \cup E^{(n)} \rangle$ , в котором есть общее множество базовых вершин  $O \subset \bigcap_{i=1}^n V^{(i)}$ . В таком графе небазовые вершины и все ребра принадлежат разным классам для разных уровней.

Реберно-чередующимся путем уровня  $i$  назовем такой реберно-простой путь в подграфе  $G^{(i)}$  графа  $G$ , в котором нет двух идущих подряд ребер одного класса.

Таким образом, граф структуры сети из [10, 12] можно назвать многоуровневым группированным графом. В данной работе будем рассматривать только один уровень – канальный.

**3.2. Метод моделирования структуры сети.** Структуру канального уровня ЛВС можно описать в виде группированного графа. В нем  $D$  – множество базовых вершин (сетевые устройства);  $P$  – множество физических портов;  $P_d$  – множество портов сетевого устройства  $d$ ;  $L^{(1)}$  – отношение соединения физического уровня: два порта  $p_1, p_2 \in P$ , ассоциированные с разными устройствами, находятся в отношении  $L^{(1)}$ , если они подключены к единой среде передачи данных.

$VID \subset \mathbb{N}_0$  – множество всех меток VLAN в сети;  $I^{(2)}$  – множество канальных интерфейсов (небазовых вершин);  $(u, v)$  – канальный интерфейс, здесь  $u \subset P_o$ ,  $v \in VID$ ;  $|u| > 1$  при агрегировании каналов; если интерфейс не участвует в VLAN, то  $v = 0$ ;  $A^{(2)}$  – отношение ассоциации канального уровня;  $L^{(2)}$  – отношение соединения канального уровня: два незаблокированных канальных интерфейса  $r_1 = (u_1, v_1)$ ,  $r_2 = (u_2, v_2)$ , ассоциированные с разными устройствами, могут находиться в отношении  $L^{(2)}$ , если существуют  $p_1 \in u_1$ ,  $p_2 \in u_2$  такие, что  $(p_1, p_2) \in L^{(1)}$ ;  $F^{(2)}$  – отношение коммутации (перенаправления) канального уровня:  $(r_1, r_2) \in F^{(2)}$ , если конфигурация устройства  $d \in D$  предусматривает возможность пересылки кадров между двумя канальными интерфейсами  $r_1, r_2 \in I_d^{(2)}$ ,  $r_1 \neq r_2$ ;  $V^{(2)} = O \cup I^{(2)}$  – множество вершин канального уровня;  $E^{(2)} = A^{(2)} \cup L^{(2)} \cup F^{(2)}$  – множество ребер канального уровня;  $G^{(2)} = \langle V^{(2)}, E^{(2)} \rangle$  – граф канального уровня;  $\hat{G}^{(2)} = \langle I^{(2)}, F^{(2)} \cup L^{(2)} \rangle$  – подграф широковещательных доменов канального уровня; широковещательные домены – это его компоненты связности; путь канального уровня – реберно-чередующийся путь в подграфе  $\hat{G}^{(2)}$ .

По стандарту IEEE 802.1D избыточные пути передачи данных блокируются для предотвращения широковещательных штормов (протокол STP), из чего вытекает следующее условие.

*Условие 1. Если в графе  $G^{(2)}$  между интерфейсами  $r_1, r_2 \in I^{(2)}$  существует путь канального уровня, то он является единственным путем канального уровня между интерфейсами  $r_1$  и  $r_2$ .*

Из определения пути канального уровня вытекает следующее свойство.

*Свойство 1. В графе  $G^{(2)}$  между двумя интерфейсами канального уровня путь канального уровня существует тогда и только тогда, когда они входят в один широковещательный домен.*

Граф канального уровня сети связный неориентированный, не содержит петель и кратных ребер, может содержать циклы. Задача автоматизации выявления структуры сети (АВСС) на канальном уровне заключается в построении графа  $G^{(2)}$  по набору исходных данных от сетевых устройств. Далее в работе будем для простоты опускать индекс (2) у обозначений, так как рассматривается только канальный уровень.

**3.3. Достижимость на канальном уровне.** В сети один интерфейс канального уровня достижим от другого, если может являться для него конечной точкой передачи данных на канальном уровне. Назовем путем достижимости канального уровня такой путь канального уровня, первое и последнее ребра которого не являются ребрами коммутации. Введем на множестве  $I$  бинарное симметричное антирефлексивное отношение

достижимости  $\leftrightarrow$  такое, что  $r_1 \leftrightarrow r_2$ , если от  $r_1$  до  $r_2$  существует путь достижимости в графе  $G$ . Будем говорить, что  $r_2$  достижим на канальном уровне от  $r_1$  (далее «К-достижим от  $r_1$ »).

Пример отношения К-достижимости приведен на рисунке 1. Устройства изображены квадратными вершинами, интерфейсы – круглыми. Ребра ассоциации показаны простыми линиями, ребра соединения и коммутации – пунктирными. К-достижимость показана точечными линиями со стрелками.

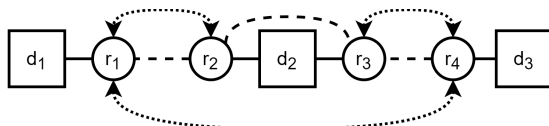


Рис. 1. Пример отношения К-достижимости

Будем говорить, что порт  $p \in P$  К-достижим от интерфейса  $r_1 \in I$ , если существует интерфейс  $(u, v) \in I$ , К-достижимый от  $r_1$ , для которого  $p \in u$ . Для каждого  $r = (u, v) \in I$  введем множество К-достижимости  $RS_r \subset I$ , содержащее все интерфейсы, К-достижимые от  $r$ . Также введем множество  $CRS_r = \bigcup_{a: a F r} RS_a$  (где  $a \in I$ ) и множество  $\leftrightarrow_r \subset \leftrightarrow$ , где  $\forall (a, b) \in \leftrightarrow_r, a = r$ .

Большинство источников исходных данных о структуре канального уровня ЛВС предоставляют информацию об отношениях соединения в неявном виде, как К-достижимость портов от интерфейсов. По определению К-достижимости и свойству 1 для К-достижимости порта  $p$  от интерфейса  $r_1$  необходимо вхождение интерфейса  $r_1 = (u_1, v_1)$  и какого-либо  $r_2 = (u_2, v_2)$  (где  $p \in u_2$ ) в один широковещательный домен. При отсутствии данных о вхождении К-достижимого порта в VLAN будем считать, что либо  $v_1 = v_2$  (если известно, что устройство поддерживает VLAN), либо  $v_2 = 0$ .

Относительно множеств  $RS$  и  $CRS$  справедлива [12] теорема:

*Теорема 1. Если канальный интерфейс  $r_2$  К-достижим от интерфейса  $r_1$ , то от  $r_1$  К-достижимы все интерфейсы, К-достижимые от интерфейсов, коммутирующих с  $r_2$ , то есть  $CRS_{r_2} \subset RS_{r_1}$ .*

Получаем свойство отношения К-достижимости, которое назовем К-транзитивностью: если  $r_1 \leftrightarrow r_2$ ,  $(r_2, r_3) \in F$  и  $r_3 \leftrightarrow r_4$ , то  $r_1 \leftrightarrow r_4$ .

Определим понятие неполноты данных о К-достижимости. Пусть  $\leftrightarrow$  – это отношение К-достижимости для графа  $G$ . Пусть  $\leftrightarrow' \subset \leftrightarrow$  – это отношение К-достижимости, построенное по исходным данным о структуре сети. Назовем данные о К-достижимости полными, если  $\leftrightarrow' = \leftrightarrow$ . Аналогично можем назвать множество К-достижимости

$RS_r$  интерфейса  $r \in I$  полным, если оно содержит все интерфейсы, К-достижимые от  $r$ . Степенью полноты данных о К-достижимости назовем число  $C(\leftrightarrow') = \frac{||\leftrightarrow' - \leftrightarrow||}{|\leftrightarrow|} \cdot 100\%$ .

При решении задачи АВСС граф  $G$  неизвестен. Имеется граф  $G'$ , полученный по исходным данным о структуре сети, в котором множества ребер канального уровня не обязательно соответствуют исходным данным о К-достижимости. Также имеется исходное отношение  $\leftrightarrow'$ , которое построено по исходным данным о К-достижимости в графе  $G'$ , извлеченным из сетевых устройств. Данные о К-достижимости обычно неполны ввиду отсутствия сведений о К-достижимости каких-либо интерфейсов в исходных данных. Исходные данные о К-достижимости могут не соответствовать ребрам графа  $G'$ , то есть элементу  $(r_i, r_j) \in \leftrightarrow'$  может не соответствовать путь К-достижимости в графе  $G$  в связи с отсутствием каких-либо ребер. Подзадача дополнения данных о К-достижимости (далее задача ДДКД) при решении задачи АВСС заключается в построении отношения  $\leftrightarrow$ , имея  $G'$  и  $\leftrightarrow'$ .

Кроме того, согласно введенным определениям задача АВСС построения графа  $G$  канального уровня заключается в построении транзитивного сокращения графа К-достижимости  $R = \langle I, \leftrightarrow \rangle$  при условии существования в графах всех вершин и ребер  $F$ . Задача АВСС может иметь несколько решений, если исходные данные неполны и в некоторых частных случаях становится NP-полной [5]. Это приводит к необходимости разработки эвристических методов решения задачи.

#### 4. Дополнение данных о достижимости на канальном уровне.

Решение задачи ДДКД является одним из подготовительных этапов алгоритма выявления структуры сети, приведенного в [10]. Оно заключается в сборе данных о достижимости от устройств, приведении этих разнородных данных к введенным понятиям модели и непосредственно дополнение данных [9]. Входными данными при решении задачи ДДКД является неполный граф  $G$  с вершинами, ребрами ассоциации и некоторыми ребрами коммутации, а также наборы сырых данных, полученные из МИБ (Management Information Base) сетевых устройств по протоколу SNMP.

**4.1. Первичная обработка исходных данных.** Доступные источники данных о связях устройств на канальном уровне включают МИБ протоколы CDP, LLDP, STP, ARP и таблицы коммутации. STP предоставляет данные только о прямых соединениях, включает данные о VLAN и заблокированных соединениях, поэтому является наиболее полным источником (достижимость интерфейсов от интерфейсов); CDP и LLDP предоставляют данные о прямых соединениях, но не включают данные о VLAN (в общем случае) и заблокированных соединениях (достижи-



мость портов от портов); AFT включают данные о К-достижимости с учетом VLAN (достижимость портов от интерфейсов); ARP – данные о К-достижимости без учета VLAN (достижимость портов от портов).

По SNMP получить эти данные можно из BRIDGE-MIB, Q-BRIDGE-MIB, CISCO-CDP-MIB, LLDP-MIB, IP-MIB, RFC1213-MIB.

В итоге переноса данных на понятия модели (реализацию опустим) для каждого канального интерфейса  $r \in I$  в графе будет получено множество К-достижимости  $RS_r$ , возможно, пустое. Также будет построено начальное отношение К-достижимости  $\leftrightarrow'$ .

Как отмечено выше, источники данных, используемые для наполнения множеств К-достижимости, не должны рассматриваться как равные. Для разделения сведений о К-достижимости введем понятия приоритетов записей о К-достижимости [10]. Введем отображение  $PR: \leftrightarrow \rightarrow \mathbb{N}$  по следующему правилу, в зависимости от источника данных о К-достижимости:

$$PR((r_1, r_2)) = \begin{cases} 0, & \text{источник – правила дополнения множеств RS;} \\ 1, & \text{источник – ARP;} \\ 2, & \text{источник – AFT;} \\ 3, & \text{источник – LLDP;} \\ 4, & \text{источник – CDP;} \\ 5, & \text{источник – STP.} \end{cases}$$

Отображение  $PR$  назовем приоритетом К-достижимости. Знание приоритета позволит общим образом различать источники данных о К-достижимости и использовать это знание при разрешении неопределенностей.

При наполнении множеств К-достижимости, если данные о К-достижимости между двумя интерфейсами обнаружены в нескольких источниках, будем выбирать источник с максимальным приоритетом.

Из-за возможной неполноты данных  $\leftrightarrow'$  необходимо подготовить для дальнейшего использования в алгоритмах таким образом, чтобы его элементы не противоречили свойству симметричности отношения К-достижимости с помощью процедуры П1 в листинге 1:

```

for  $r_i \in I$ :
  for  $r_j : (r_i, r_j) \in \leftrightarrow'$ 
    if  $(r_j, r_i) \notin \leftrightarrow'$ :
       $\leftrightarrow'.add((r_j, r_i))$ 

```

Листинг 1. Процедура П1 подготовки множеств К-достижимости

Вычислительная сложность процедуры П1 составляет  $O(|I|^2)$ .

**4.2. Получение точного решения при ограничениях.** Используя свойства отношения К-достижимости, приведем численный метод построения более точного графа К-достижимости – решения задачи ДДКД – из [9] с поправкой на свойства и отношения, введенные в настоящей работе. В основе метода – построение замыкания отношения  $\leftrightarrow'$  относительно свойства К-транзитивности. Метод представлен в виде алгоритма А1 в листинге 2:

```

for  $r_i \in I$ :
   $Q = \text{Queue}; \text{Visited} = \{r_i\};$ 
  for  $(r_i, r_j) \in \leftrightarrow'$ :
     $Q.\text{enqueue}((r_j, 1))$ 
  while  $Q \neq \emptyset$ :
     $(r_j, q) = Q.\text{dequeue}()$ 
     $\text{Visited}.\text{add}(r_j)$ 
    if  $q = 1$ :
      if  $(r_j, r_i) \notin \leftrightarrow'$ :
         $\leftrightarrow'.\text{add}((r_j, r_i))$ 
      if  $(r_i, r_j) \notin \leftrightarrow'$ :
         $\leftrightarrow'.\text{add}((r_i, r_j))$ 
      for  $r_k : (r_j, r_k) \in F$ :
        if  $r_k \notin \text{Visited}$ :
           $Q.\text{enqueue}((r_k, 0))$ 
    if  $q = 0$ :
      for  $(r_j, r_k) \in \leftrightarrow'$ :
         $Q.\text{enqueue}((r_k, 1))$ 

```

Листинг 2. Алгоритм А1 решения задачи ДДКД

В [9] аналогичный алгоритм полагался на обработку матрицы отношения К-достижимости, в связи с чем вычислительная сложность была определена как  $O(|I|^3)$ . Вычислительная сложность приведенного в листинге 2 метода составляет  $O(|I|^2)$ : для каждого интерфейса канального уровня производится операция, аналогичная обходу в ширину графа К-достижимости  $R$ , что эквивалентно  $O(|I|)$  при использовании эффективных структур данных. Для определения границ применимости метода в [9] доказана:

*Теорема 2.* Пусть интерфейсы  $I_i$  и  $I_j$  находятся в отношении  $L$  друг с другом в графе  $G$ . Если изначально  $(I_i, I_j) \notin \leftrightarrow'$ , то Алгоритм А1 не позволяет определить К-достижимы ли указанные интерфейсы друг от друга.

На основе свойств К-достижимости и теоремы 2 можно вывести условие возможности построения отношения  $\leftrightarrow$  с полными данными (единственного решения задачи ДДКД) с помощью Алгоритма А1:

*Следствие 1.* Если для любых двух интерфейсов  $I_i$  и  $I_j$ , находящихся в отношении  $L$ , изначально  $(I_i, I_j) \in \leftrightarrow'$ , тогда с помощью Алгоритма А1 возможно построить единственное отношение  $\leftrightarrow$  с полными данными о  $K$ -достижимости.

Таким образом, есть прямая зависимость полноты множеств  $K$ -достижимости, получаемых алгоритмом А1, от изначальных сведений о  $K$ -достижимости соединенных интерфейсов. И решение задачи ДДКД сводится к построению замыкания отношения  $\leftrightarrow'$  относительно свойства  $K$ -транзитивности.

Главным предположением для использования приведенного метода является отсутствие в исходных данных некорректной информации о  $K$ -достижимости ( $K$ -достижимость интерфейсов, на самом деле недостижимых) и наличие всех ребер коммутации. В зависимости от качества исходных данных в реальных условиях можно получить или отношение  $\leftrightarrow$  с полными данными о  $K$ -достижимости, или отношение  $\leftrightarrow''$  такое, что  $\leftrightarrow' \subset \leftrightarrow'' \subset \leftrightarrow$ .

**4.3. Использование достижимости при выявлении структуры сети.** Полученное отношение  $K$ -достижимости возможно использовать для построения отношений  $L$  и  $F$  на канальном уровне. Для этого в [10] приведены следующие правила:

*Правило 1.* Если от интерфейса  $r_1 \in I$   $K$ -достижим только один интерфейс  $r_2 \in I$  и  $RS_{r_2} = CRS_{r_1} \cup \{r_1\}$ , то  $(r_1, r_2) \in L$ .

*Правило 2.* Если два интерфейса  $r_1$  и  $r_2$   $K$ -достижимы друг от друга, и  $RS_{r_1} = CRS_{r_2} \cup \{r_2\}$  и  $RS_{r_2} = CRS_{r_1} \cup \{r_1\}$ , то  $(r_1, r_2) \in L$ .

*Правило 3.* Если для устройства  $d \in D$  и двух  $K$ -достижимых друг от друга интерфейсов канального уровня  $r_1$  и  $r_2$  выполняется:

1. От  $r_1$   $K$ -достижим  $r_3 \in I_d$ , от  $r_2$   $K$ -достижим  $r_4 \in I_d$ .

2. От  $r_1$  и  $r_2$  не  $K$ -достижимы интерфейсы из  $I_d$ , отличные от  $r_3$  и  $r_4$  соответственно.

3. От интерфейсов, коммутирующих с  $r_1$  или  $r_2$ , не  $K$ -достижимы интерфейсы, ассоциированные с  $d$ .

Тогда  $r_3$  и  $r_4$  находятся в отношении  $F$ .

Алгоритм А2 [10] построения этих отношений в графе  $G$  заключается в следующем:

1. Дополнение данных о  $K$ -достижимости.
2. Для пар интерфейсов  $I$  проверка правил 1 и 2 и добавление пары в  $L$  при выполнении правила.
3. Для пар интерфейсов одного устройства проверка правила 3 и добавление пары в  $F$  при выполнении правила.

4. Возврат к шагу 1 если было обнаружено хотя бы одно ребро  $L$  или  $F$ .

Алгоритм отчасти заключается в построении транзитивного сокращения графа  $R$ . При использовании эффективных структур данных и алгоритма A1 вычислительная сложность алгоритма A2 составит  $O(I^5)$  и  $\Omega(I^3)$  против  $O(I^8)$  и  $\Omega(I^4)$ , показанных в [10].

**4.4. Методы выбора решения при неоднозначности.** Условие из Следствия 1 легко достигается в реальных сетях при использовании всех доступных источников данных о достижимости. Однако отсутствие данных о К-достижимости соединенных интерфейсов все же возможно. Пример такой ситуации приведен на рисунке 2. Из отношения К-достижимости видно, что граф  $G$  связный, но правила из раздела 4.3 не позволяют обнаружить соединения  $L$ , кроме  $(r_6, r_7)$ .

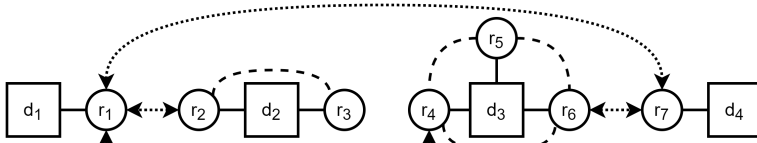


Рис. 2. Пример неоднозначности К-достижимости

Из Следствия 1 очевидно, что отсутствовать в отношении достижимости в первую очередь должны пары интерфейсов, находящиеся в отношении  $L$  в графе  $G$ . Исходя из этого построим отношение потенциальной К-достижимости  $\leftrightarrow^*$  по следующему правилу: интерфейсы  $(r_1, r_2) \in \leftrightarrow^*$ , если все следующие условия выполняются:

- $(r_1, r_2) \notin \leftrightarrow$ ;
- $r_1, r_2$  не находятся в отношении  $L$  с другими интерфейсами;
- $r_1 \in I_{d_1}, r_2 \in I_{d_2}$  и  $d_1 \neq d_2$ ;
- $RS_{r_1} \cap RS_{r_2} = \emptyset$ ;
- $CRS_{r_1} \cap CRS_{r_2} = \emptyset$ ;
- $RS_{r_1} \cap CRS_{r_2} \neq \emptyset$  или  $RS_{r_2} \cap CRS_{r_1} \neq \emptyset$ ;
- $r_1 \notin CRS_{r_2}$  и  $r_2 \notin CRS_{r_1}$ .

Таким образом, потенциально К-достижимы пары интерфейсов, которые не противоречат правилу 2 и другим свойствам К-достижимости, а также находятся в одном широковещательном домене.

Использовать отношение потенциальной К-достижимости в алгоритме A2 можно следующим образом. После шага 4, если отношение  $\leftrightarrow^*$  не пусто, выбрать одну из пар  $(r_1, r_2) \in \leftrightarrow^*$  и перенести в  $\leftrightarrow$  и  $L$ , после чего продолжить с шага 1. Сложность алгоритма ( $O(I^5)$  и  $\Omega(I^3)$ ) при этом

не возрастает, так как условие возврата на новую итерацию аналогично условию из шага 4.

В примере на рисунке 2 потенциально  $K$ -достижимы пары интерфейсов  $(r_3, r_4)$  и  $(r_2, r_5)$ . Используя  $\leftrightarrow^*$  и модифицированный алгоритм A2, можно получить два решения. Они приведены на рисунке 3.

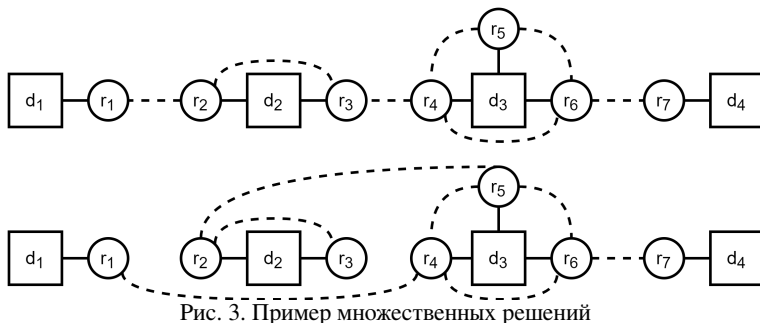


Рис. 3. Пример множественных решений

Неоднозначность решения можно разрешить в некоторых ситуациях с помощью приоритета  $K$ -достижимости  $PR$ . Заметим, что некоторые источники данных о связях устройств предоставляют сведения только о соединениях напрямую (LLDP, CDP, STP). Введем число  $Z$ , которое определяет границу в приоритете  $K$ -достижимости между данными о прямых соединениях и прочими данными. Для описанного отображения  $PR$ ,  $Z = 3$ . Располагая такими дополнительными сведениями, будем исключать элементы  $(r_1, r_2)$  из отношения  $\leftrightarrow^*$ , если существует  $q \in \leftrightarrow^*_{r_1} \cup \leftrightarrow^*_{r_2}$ , для которого  $PR(q) \geq Z$ .

Таким образом, при невыполнении условий из следствия 1 возможно существование нескольких решений задачи выявления структуры сети, но располагая дополнительными сведениями, можно выбрать из множества решений наиболее точное.

**4.5. Исключение некорректных данных.** Исходные данные о достижимости могут быть некорректны, то есть указывать на  $K$ -достижимость интерфейсов, на самом деле недостижимых. Происходить это может из-за разнородности сетевого оборудования и в высокодинамичных сетях. Например, беспроводное устройство может быстро переключиться между точками доступа в разных VLAN и остаться в кэше протокола ARP и в таблицах коммутации разных устройств на разных интерфейсах.

Условие обнаружения некорректных данных – невыполнение условия 1, что можно обнаружить через отношение  $K$ -достижимости:

*Следствие 2.* Для  $r_1, r_2 \in I_d$ ,  $d \in D$ , таких что  $(r_1, r_2) \in F$ , множества  $RS_{r_1}$  и  $RS_{r_2}$  не должны пересекаться.

Пересечение множеств  $RS$  говорит о существовании цикла в структуре канального уровня, что невозможно в функционирующей сети. Если они пересекаются, с точностью определить, какой из потенциальных путей в цикле на самом деле существует, невозможно.

Продолжить выявление структуры сети при некорректных данных возможно, исключив из рассмотрения некоторые элементы, образующие цикл. При обнаружении некорректных данных выбирается один из интерфейсов  $r_i$  пары  $(r_1, r_2) \in F$ , от которой обнаружен цикл, и из отношения  $\leftrightarrow$  исключаются все элементы  $(r_i, r_j)$ , где  $r_j \in RS_{r_1} \cap RS_{r_2}$ . Таким образом, цикл будет разорван, и устройство, из-за которого образовался цикл, окажется с одной стороны пары  $(r_1, r_2)$ .

При реализации данного метода следует принимать во внимание возможность существования циклов с одними и теми же устройствами и портами в разных ширококвещательных доменах. При разрешении некорректности следует выбирать одинаковое направление разрыва циклов, чтобы соблюсти физическую связность сети. Также предпочтение в выборе опорной пары интерфейсов следует отдавать интерфейсам устройств, которые доступны по SNMP, так как для таких устройств данные более достоверны.

**5. Вычислительные испытания.** Приведенные в работе методы были реализованы в рамках программной системы для автоматизированного выявления структуры сети, предыдущие реализации которой описаны в [10].

**5.1. Испытания со сгенерированными данными.** Реализация алгоритмов протестирована на 1000 сгенерированных сетей различных размеров: 1–3 маршрутизатора, 2–50 коммутаторов, 40–1000 рабочих станций, 1–5 VLAN, 85–2000 интерфейсов. Сначала производится генерация сети и построение правильного отношения достижимости  $\leftrightarrow$ . Затем генерируется отношение  $\leftrightarrow'$  различной полноты. После этого производится дополнение данных о К-достижимости с помощью разработанных алгоритмов и сравнение полученного отношения К-достижимости с правильным, а также построение графа  $G$  структуры сети и сравнение его со сгенерированным графом. Для построения графа  $G$  из сгенерированного графа исключались все ребра  $L$ .

При любых размерах изначального графа и любой полноте данных о К-достижимости и выполнении условий из следствий 1 и 2 с помощью реализации алгоритмов удалось получить отношение К-достижимости с полными данными, а построенные графы  $G$  совпадали со сгенери-

рованным графом. Отметим, что данные, с помощью которых можно удовлетворить условия из следствия 1, составляют 1–5% от данных о К-достижимости (степень полноты  $C(\leftrightarrow')$ ), то есть с помощью разработанных методов удастся дополнить 95–99% данных о К-достижимости.

Проведены испытания реализации алгоритмов в случаях, когда условия из следствия 1 не выполняются. Используя этот же метод генерации (и сети размером 40–130 устройств и 1–5 VLAN), из отношения  $\leftrightarrow'$  исключались случайно выбранные пары интерфейсов, соединенные ребром  $L$  в сгенерированном графе. Степень полноты данных о непрямых соединениях варьировалась от 10 до 90%. Все данные о К-достижимости приняты с одинаковым приоритетом.

Введем множество  $\leftrightarrow^- = \leftrightarrow' \setminus \leftrightarrow$ , которое содержит пары К-достижимых интерфейсов, некорректно присутствующих в дополненном отношении  $\leftrightarrow'$ . Введем число  $C^-(\leftrightarrow') = \frac{||\leftrightarrow^-|| + ||\leftrightarrow' - \leftrightarrow^-||}{||\leftrightarrow||} \cdot 100\%$ , которое показывает степень неполноты данных о К-достижимости с учетом неправильно дополненных данных.

Данные тестирования приведены в таблице 1. Столбец 2 содержит количество пар интерфейсов в правильном отношении К-достижимости. Столбцы 3 и 4 содержат среднюю степень полноты данных  $C(\leftrightarrow')$  и  $C^-(\leftrightarrow')$  в полученном отношении К-достижимости с учетом и без учета потенциальной К-достижимости. Столбцы 5 и 6 содержат максимальное и среднее число пар К-достижимых интерфейсов, которые алгоритм не смог обнаружить (с учетом потенциальной К-достижимости).

Таблица 1. Результаты тестирования при неоднозначности К-достижимости

Кол-во VLAN	Кол-во отсутствующих соединений	Ср. кол-во пар в $\leftrightarrow$	Средняя $C(\leftrightarrow')$ после обработки, %	Ср. $C^-(\leftrightarrow')$ после обработки с учетом $\leftrightarrow^*$ , %	Макс. ненайденных пар	Ср. ненайденных пар
1–5	0	5157	100	100	0	0
1–2	1	6958	99,20	99,93	66	2,69
1–2	2	6958	98,61	99,81	100	6,98
1–2	3	6958	98,36	99,68	96	11,02
1–2	4	6958	97,51	99,39	218	20,40
1–2	5	6958	97,04	99,26	214	26,27
3–5	1	3957	99,40	99,76	48	8,83
3–5	2	3957	98,71	99,63	122	10,77
3–5	3	3957	98,11	99,46	180	15,30
3–5	4	3957	97,68	99,28	280	20,75
3–5	5	3957	96,99	99,06	222	24,41

Согласно результатам тестирования приведенные методы способны построить отношение К-достижимости с данными, близкими к полным, даже при отсутствии части критических для точного решения задачи данных. Использование отношения потенциальной К-достижимости позволяет улучшить результат на 0,5–2%.

Полученный результат можно улучшить еще, если использовать данные о приоритете К-достижимости. Возможно уменьшить (в зависимости от конфигурации сети) вероятность ошибочного выбора из потенциально К-достижимых элементов, что позволит повысить точность выявления структуры сети. Тем не менее использование потенциальной К-достижимости позволит получить связный граф структуры сети, близкий к правильному, даже при отсутствии сведений о приоритете.

Время обработки реализации алгоритма дополнения данных с учетом потенциальной К-достижимости для сетей с 300 интерфейсами канального уровня составило в среднем 647 мс.

**5.2. Испытания с реальными данными.** Испытания реализации методов протестировано в реальной ЛВС Петрозаводского государственного университета. Исследуемый сегмент сети содержит: 1 маршрутизатор Cisco 7600, 4 коммутатора 3 уровня Cisco 3750 и Cisco 3850, 52 различных коммутаторов Cisco 2 уровня доступные по SNMP и 296 недоступных коммутаторов, точек доступа Wi-Fi и прочих служебных устройств, 795 рабочих станций и 101 VLAN (всего 1148 устройств). Количество интерфейсов канального уровня составляет 6037. Время сбора данных составило 11 минут 23 секунды, построения описания – 2 минуты 7 секунд.

Основные источники данных о достижимости – CDP, STP и таблицы коммутации. Отношение К-достижимости для сравнения было построено с помощью данных, полученных от сетевых администраторов и самостоятельно из сети. Используя разработанные алгоритмы, удалось построить отношение К-достижимости с полными данными и точный граф структуры сети *G*. Не было сделано ни одного предположения о потенциальной К-достижимости, не обнаружено некорректных данных.

Всего было обнаружено 99270 пар К-достижимых интерфейсов. Дополненная алгоритмами часть составила 90%. Такое большое число объясняется большим количеством устройств, VLAN и недоступных по SNMP служебных устройств – много данных было недоступно изначально, однако кэши CDP и STP предоставили большую часть необходимых данных о напрямую соединенных интерфейсах.

Также испытания были проведены в сети филиала АО «АЭМ-технологии» «Петрозаводскмаш». Обнаружено 427 устройств (Cisco, HP, Telesis), 1325 портов, 441 соединение на физическом уровне, 9079 интер-



фейсов канального уровня, 2455 соединений на канальном уровне, 3579 связей коммутации. Время сбора данных составило 19 минут 44 секунды, построения описания – 6 минут 20 секунд. Было сделано 4 предположения о потенциальной достижимости и обработано 17 случаев некорректности данных о достижимости. Удалось построить точный граф структуры сети (подтверждается существующей документацией), кроме 6 отсутствующих соединений с хостами (отсечены в ходе обработки некорректных данных).

**5.3. Сравнение с аналогами.** Предыдущий вариант разработанных методов был сравнен в [9] с результатами работ других исследователей [5–8]. Было показано существенное преимущество в точности (95-99% дополненных данных против 50%) при повышении вычислительной сложности.

Приведены методы дополнения данных в случаях их некорректности или неоднозначности, которые позволяют улучшить предыдущий результат еще на 1-1,5% без повышения вычислительной сложности.

Использование некорректных данных не было исследовано ни в одной из известных авторам предыдущих работ. Использование неоднозначных данных было исследовано в [5], где предлагалось производить построение ребер соединения с помощью аналогичного отношения потенциальной  $K$ -достижимости. Однако в указанной работе предлагалось сразу строить граф  $G$  от потенциальной  $K$ -достижимости, минуя стадию построения ребер  $L$ , с точностью известных по  $K$ -достижимости. Это приводило к большому количеству ошибок в построенном графе. Также в работе не использовались приоритеты  $K$ -достижимости для повышения точности предположений. Кроме того, предложенные в [5] алгоритмы не совместимы с сетями с VLAN.

**6. Заключение.** Описание физической и логической структур локальной вычислительной сети необходимо для решения многих задач сетевого управления. Однако автоматизация такого описания осложнена возможностью неполноты и некорректности исходных данных о соединениях сетевых устройств.

Представлены модели и методы обработки неполных исходных данных о достижимости устройств на канальном уровне сети. Дополнение таких данных позволяет выявить элементы структуры сети на физическом и канальном уровнях. В работе сформулирована и проанализирована задача дополнения неполных данных, показана ее близость с классической задачей построения замыкания бинарного отношения. Определено свойство, относительно которого строится замыкание, предложен простой алгоритм решения задачи, основанный на обходе графа достижимости на канальном уровне в ширину. Определено условие для исходных данных о

достижимости, при выполнении которого возможно получить точное решение задачи: известность достижимости устройств, соединенных между собой физически.

Определены и проанализированы случаи, когда получение точного решения не гарантируется, в том числе, когда исходные данные некорректны. Для них предложены и обоснованы методы исключения некорректных данных и выбора одного решения из множества. В этих методах используются свойства модели структуры сети и дополнительные сведения об исходных данных (из какого источника они получены).

Представленные результаты вычислительных испытаний (в том числе в реальных сетях) показывают применимость и эффективность разработанных методов при различных конфигурациях ЛВС. Удается получить точное решение при выполнении обозначенного условия для исходных данных и решение, близкое к точному (98-100% полноты данных о достижимости после обработки), и невыполнении или некорректности исходных данных. Апробация, проведенная в двух крупных вычислительных сетях реальных предприятий, показывает возможность возникновения в реальных условиях неполноты и некорректности данных, которую разработанные методы успешно смогли исправить.

Авторы выражают благодарность д. т. н., профессору кафедры ПМИК ПетрГУ Р. В. Воронову за консультацию в области теории графов.

### Литература

1. *Hussain T. H., Habib S. J.* Capacity planning of network redesign — A case study // Proceedings of the 2010 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'10). 2010. pp. 52–57.
2. *Zhu Z. et al.* Data flow monitoring and control of LAN based on strategy // 2010 International Conference on Networking and Digital Society. 2010. vol. 2. pp. 225–228.
3. *Sivakumar L. et al.* Virtual topologies for abstraction service for IP-VPNs // 2016 17th International Telecommunications Network Strategy and Planning Symposium (Networks). 2016. pp. 213–220.
4. *Wang C. et al.* A method of network topology optimization design considering application process characteristic // Modern Physics Letters B. 2018. vol. 32. no. 07. pp. 1850091–1–1850091-11.
5. *Gobjuka H., Breitbart Y. J.* Ethernet topology discovery for networks with incomplete information // IEEE/ACM Transactions on Networking. 2010. vol. 18. no. 4. pp. 1220–1233.
6. *Zichao L. et al.* Ethernet topology discovery for virtual local area networks with incomplete information // 2014 4th IEEE International Conference on Network Infrastructure and Digital Content. 2014. pp. 252–256.
7. *Bejerano Y.* Taking the skeletons out of the closets: a simple and efficient topology discovery scheme for large ethernet LANs // IEEE/ACM Transactions on Networking. 2009. vol. 17. no. 5. pp. 1385–1398.

8. Zhou J., Ma Y. Topology discovery algorithm for ethernet networks with incomplete information based on VLAN // 2016 IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC). 2016. pp. 396–400.
9. Andreev A., Shabaev A., Bogoiavlenskii I. Network Topology Discovery: a Problem of Incomplete Data Improvement // 2019 24th Conference of Open Innovations Association (FRUCT). 2019. pp. 10–16.
10. Andreev A., Bogoiavlenskii I. An algorithm for building an enterprise network topology using widespread data sources // 2017 21st Conference of Open Innovations Association (FRUCT). 2017. pp. 34–43.
11. Андреев А. А. Обобщенная графовая модель виртуальных частных сетей в коммуникационной инфраструктуре локального поставщика сетевых услуг // Программная инженерия. 2017. Т. 8. № 6. С. 243–249.
12. Андреев А. А. и др. Обобщенная графовая модель структуры физического, канального и сетевого уровней ИКТ-инфраструктуры локального поставщика сетевых услуг // Программная инженерия. 2016. Т. 7. № 9. С. 400–407.
13. Sun Y., Shi Z., Wu Z. A discovery algorithm for physical topology in switched ethernets // The IEEE Conference on Local Computer Networks 30th Anniversary (LCN'05) 1. 2005. pp. 311–317.
14. Zhangchao W. et al. An algorithm and implementation of network topology discovery based on SNMP // 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI). 2016. pp. 283–286.
15. Zhou S. et al. Research on Network Topology Discovery Algorithm for Internet of Things Based on Multi-Protocol // 2018 10th International Conference on Modelling, Identification and Control (ICMIC). 2018. pp. 1–6.
16. Ma X., Yu T. An algorithm of physical network topology discovery in multi-VLANs // *Telkomnika*. 2016. vol. 14. no. 3A. pp. 375–379.
17. Jiang J., Xu X. L., Cao N. Research on improved physical topology discovery based on SNMP // 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). 2017. vol. 2. pp. 219–222.
18. Zhang X. et al. An Optimization Algorithm of Network Topology Discovery Based on SNMP Protocol // *Journal of Computer and Communications*. 2018. vol. 6. no. 01. pp. 104–111.
19. Zhou Y. et al. Discovery algorithm for network topology based on SNMP // 2015 International Conference on Automation, Mechanical Control and Computational Engineering. 2015. pp. 1623–1628.
20. Li D. et al. Research and Application of Heterogeneous Network Topology Discovery Algorithm Based on Multiple Spanning Tree Protocol // *Knowledge Discovery and Data Mining*. 2012. pp. 437–444.

**Андреев Антон Александрович** – аспирант, Петрозаводский государственный университет. Область научных интересов: применение методов математического моделирования в компьютерных сетях. Число научных публикаций – 20. andreev@cs.petrstu.ru; пр. Ленина, 33, 185910, Петрозаводск, Россия; р.т.: +7(814)271-10-01, факс: +7(814)271-10-00.

**Шабает Антон Игоревич** – канд. техн. наук, доцент, кафедра прикладной математики и кибернетики, Петрозаводский государственный университет. Область научных интересов: математическое моделирование систем управления, методы решения оптимизационных задач и их применение в управлении организациями и предприятиями, промышленная автоматизация. Число научных публикаций – 68. ashabaev@petrstu.ru; пр. Ленина, 31, 185910, Петрозаводск, Россия; р.т.: +7(814)271-32-10, +7(814)271-10-68.

A.A. ANDREEV, A.I. SHABAEV  
**MODELS AND METHODS FOR DISCOVERY OF LOCAL AREA  
NETWORK TOPOLOGY WITH INCOMPLETE DATA**

---

*Andreev A.A., Shabaev A.I. Models and Methods for Discovery of Local Area Network Topology with Incomplete Data.*

**Abstract.** A lot of network management tasks require a description of the logical and physical computer network topology. Obtaining such a description in an automatic way is complicated due to the possibility of incompleteness and incorrectness of the initial data on the network structure. This article provides a study on the properties of incomplete initial data on network device connectivity on the link layer. Methods for generalized handling of the heterogeneous input data on the link layer are included. We describe models and methods for deriving a missing part of the data, as well as the condition in which it is possible to get a single correct network topology description. The article includes algorithms for building a link layer topology description from incomplete data when this data is possible to fulfill up to the required level. Also, we provide methods for detecting and resolving an ambiguity in the data and methods for improving incorrect initial data. The tests and evaluations provided in the article demonstrate the applicability and effectiveness of the build methods for discovering various heterogeneous real-life networks. Additionally, we show the advantages of the provided methods over the previous analogs: our methods are able to derive up to 99% data on link layer connectivity in polynomial time; able to provide a correct solution from an ambiguous data.

**Keywords:** Computer Networks, Modeling, Automation, Network Topology, Link Layer

---

**Andreev Anton Aleksandrovich** – Postgraduate Student, Petrozavodsk State University. Research interests: applying of mathematical modeling in computer networks. The number of publications – 20. andreev@cs.petsu.ru; 33, Lenina st., 185910, Petrozavodsk, Russia; office phone: +7(814)271-10-01; fax: +7(814)271-10-00.

**Shabaev Anton Igorevich** – Ph.D., Associate Professor, Chair of Applied Mathematics and Cybernetics, Petrozavodsk State University. Research interests: mathematical modeling of control systems, methods of the optimization problems solution, industrial automation. The number of publications – 68. ashabaev@petsu.ru; 31, Lenina st., 185910, Petrozavodsk, Russia; office phone: +7(814)271-32-10, +7(814)271-10-68.

## References

1. Hussain T. H., Habib S. J. Capacity planning of network redesign — A case study. Proceedings of the 2010 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'10). 2010. pp. 52–57.
2. Zhu Z. et al. Data flow monitoring and control of LAN based on strategy. 2010 International Conference on Networking and Digital Society. 2010. vol. 2. pp. 225–228.
3. Sivakumar L. et al. Virtual topologies for abstraction service for IP-VPNs. 2016 17th International Telecommunications Network Strategy and Planning Symposium (Networks). 2016. pp. 213–220.
4. Wang C. et al. A method of network topology optimization design considering application process characteristic. *Modern Physics Letters B*. 2018. vol. 32. no. 07. pp. 1850091-1–1850091-11.

5. Gobjuka H., Breitbart Y. J. Ethernet topology discovery for networks with incomplete information. *IEEE/ACM Transactions on Networking*. 2010. vol. 18. no. 4. pp. 1220–1233.
6. Zichao L. et al. Ethernet topology discovery for virtual local area networks with incomplete information. 2014 4th IEEE International Conference on Network Infrastructure and Digital Content. 2014. pp. 252–256.
7. Bejerano Y. Taking the skeletons out of the closets: a simple and efficient topology discovery scheme for large ethernet lans. *IEEE/ACM Transactions on Networking*. 2009. vol. 17. no. 5. pp. 1385–1398.
8. J. Zhou, Y. Ma. Topology discovery algorithm for ethernet networks with incomplete information based on VLAN. 2016 IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC). 2016. pp. 396–400.
9. Andreev A., Shabaev A., Bogoiavlenskii I. Network Topology Discovery: a Problem of Incomplete Data Improvement. 2019 24th Conference of Open Innovations Association (FRUCT). 2019. pp. 10–16.
10. Andreev A., Bogoiavlenskii I. An algorithm for building an enterprise network topology using widespread data sources. 2017 21st Conference of Open Innovations Association (FRUCT). 2017. pp. 34–43.
11. Andreev A. A. [Generalized Graph Model of Virtual Private Networks in the Communication Infrastructure of a Local Network Service Provider]. *Programmnyaya Ingeneria — Software Engineering*. 2017. Issue 8. vol. 6. pp. 243–249. (In Russ.).
12. Andreev A. A. et al. [Generalized Graph Model of the Physical, Link and Network Layers Structure of the ICT-Infrastructure of a Local Network Service Provider]. *Programmnyaya Ingeneria — Software Engineering*. 2016. Issue 7. vol. 9. pp. 400–407. (In Russ.).
13. Sun Y., Shi Z., Wu Z. A discovery algorithm for physical topology in switched ethernet. The IEEE Conference on Local Computer Networks 30th Anniversary (LCN'05) I. 2005. pp. 311–317.
14. Zhangchao W. et al. An algorithm and implementation of network topology discovery based on SNMP. 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI). 2016. pp. 283–286.
15. Zhou S. et al. Research on Network Topology Discovery Algorithm for Internet of Things Based on Multi-Protocol. 2018 10th International Conference on Modelling, Identification and Control (ICMIC). 2018. pp. 1–6.
16. Ma X., Yu T. An algorithm of physical network topology discovery in multi-VLANs. *Telkommika*. 2016. vol. 14. no. 3A. pp. 375–379.
17. Jiang J., Xu X. L., Cao N. Research on improved physical topology discovery based on SNMP. 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). 2017. vol. 2. pp. 219–222.
18. Zhang X. et al. An Optimization Algorithm of Network Topology Discovery Based on SNMP Protocol. *Journal of Computer and Communications*. 2018. vol. 6. no. 01. pp. 104–111.
19. Zhou Y. et al. Discovery algorithm for network topology based on SNMP. 2015 International Conference on Automation, Mechanical Control and Computational Engineering. 2015. pp. 1623–1628
20. Li D. et al. Research and Application of Heterogeneous Network Topology Discovery Algorithm Based on Multiple Spanning Tree Protocol. *Knowledge Discovery and Data Mining*. 2012. pp. 437–444.