

Практические аспекты использования графов знаний для моделирования телекоммуникационных сетей

Н.А. Жукова
Санкт-Петербургский институт
информатики и автоматизации
Российской академии наук
Санкт-Петербург, Россия
ORCID: 0000-0001-5877-4461
nazhukova@mail.ru

И.А. Куликов
Кафедра МО ЭВМ
Санкт-Петербургский
государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
Санкт-Петербург, Россия
ORCID: 0000-0002-2532-5579
i.a.kulikov@gmail.com

Аннотация— Доклад посвящен практическим аспектам использования графов при моделировании телекоммуникационных сетей. Представлен обзор применяемых графовых моделей сетей, приведены подлежащие моделированию характеристики сетей, показаны преимущества и недостатки используемых графовых моделей. Проанализированы требования к моделям телекоммуникационных сетей — возможность динамического управления параметрами сети, доступом к информации, уровнем качества предоставляемых сетями услуг конечным пользователям. Представлены оценки соответствия применяемых графовых моделей современным и перспективным требованиям операторов. Приведено определение графа знаний и проанализированы перспективы практического использования моделей на основе графов знаний при моделировании телекоммуникационных сетей. Сформулированы задачи, которые можно решать при помощи моделей, основанных на графах знаний. На практическом примере показана возможность решать задачи, стоящие перед операторами, при помощи моделей, построенных на основе графов знаний.

Ключевые слова: телекоммуникационная сеть, граф знаний, моделирование

I. ВВЕДЕНИЕ

В настоящее время для моделирования телекоммуникационных сетей используются различные модели – графовые, вероятностные, модели массового обслуживания и пр. При этом, как сами модели, так и информационные системы, построенные на их основе, оперируют только теми данными, которые необходимы для их функционирования. Когда возникает необходимость в обработке данных, содержащихся в разных информационных системах, приходится строить обобщенные модели, объединяющие ряд уже существующих моделей. Проблема является актуальной и часто встречается на практике, например, при необходимости объединения бизнес-профиля абонента сети и данных о его поведении (статистика использования сервисов и приложений, предлагаемых сетью, геоданные и пр.). Для построения таких обобщенных моделей могут быть использованы различные подходы. В данной статье авторами предлагается решение на основе графа знаний. Представлены предпосылки для использования графа знаний, описана сама технология, применяемые онтологии и пример построения графа знаний

телекоммуникационной сети, решающий практическую задачу. Показаны преимущества использования графа знаний как технологии для построения обобщенной модели телекоммуникационной сети.

II. ОБЗОР СУЩЕСТВУЮЩИХ РЕШЕНИЙ

Схожие задачи обработки данных телекоммуникационных сетей решаются следующими классами систем:

- Системы обработки машинных данных (например, Splunk [8], Datadog [9], Logstash [10]).
- Аналитические системы (например, IBM Telecom Analytics Solutions [11], Huawei Big Data Analytics (BDA) [12]).

Системы обработки машинных данных обрабатывают события, логируемые сетевыми устройствами, выделяют в них выбранные оператором паттерны, агрегируют их и используют в качестве исходных данных для отчетов и для построения различных панелей управления. Объединение моделей происходит на самом верхнем уровне анализа данных, когда первичные данные уже обработаны и часть взаимосвязей может быть потеряна. Для добавления новых взаимосвязей на уровне первичных данных требуется перепроектирование на уровне паттернов логов, агрегации данных и бизнес-логики.

Аналитические системы работают с уже подготовленными данными. Как правило, эти данные поступают из различных систем, функционирующих в сети оператора. Для добавления новых взаимосвязей на уровне первичных данных требуется вносить изменения на уровне формирования данных в системах – источниках данных, на уровне обработчиков входящих данных самой аналитической системы и вносить изменения в ее бизнес-логику.

III. СУЩЕСТВУЮЩИЕ ГРАФОВЫЕ МОДЕЛИ ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЕЙ

Графовые модели телекоммуникационных сетей широко применяются на практике, ниже приведен перечень наиболее часто встречающихся графовых моделей:

1. Графы структур сетевых слоев (устройства и каналы передачи данных, приложения, сетевая иерархия, функции, сетевые ограничения, модель данных) [1].

2. Граф атак [2].

3. Граф прав доступа [3].

4. Различные частные графовые модели для решения конкретных задач.

Традиционные графовые модели телекоммуникационных сетей успешно решают специализированные задачи, но не всегда возможно в рамках одной модели решить задачи, требующие анализа различных аспектов функционирования сети.

IV. СОСТАВ ДАННЫХ О ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЯХ ДЛЯ ПОСТРОЕНИЯ ГРАФОВЫХ МОДЕЛЕЙ

Данные для построения графов телекоммуникационных сетей включают:

- данные об устройствах и топологии сети;
- перечень сервисов;
- роли пользователей и права пользователей;
- модель данных;
- структуры пользовательских интерфейсов;
- статистика поведения пользователей по обращению к сервисам, запросам доступа к данным и фактам получения доступа;
- статистика по производительности сервисов и возникших при их выполнении ошибках;
- аналитика видов угроз и сценарии их реализации;
- лог файлы серверов и пользовательских устройств;
- журналы событий.

Перечисленные данные позволяют построить следующие графовые модели телекоммуникационной сети:

1. Модель топологии сети;
2. Модель прикладных сервисов;
3. Граф пользовательских приложений;
4. Ролевая модель;
5. Модель распространения прав доступа (граф доступа);
6. Модели пользовательских интерфейсов;
7. Модель поведения пользователя по обращению к сервисам и доступа к данным;
8. Модель атак;
9. Граф состояния сервисов (срез на момент времени).

Перечень моделей не полон и может быть расширен в зависимости от состава доступных данных.

V. ИСПОЛЬЗОВАНИЕ ГРАФОВ ЗНАНИЙ ПРИ МОДЕЛИРОВАНИИ ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЕЙ

Графы знаний используются как для открытых проектов (открытые графы знаний), так и для корпоративных — индустриальные графы знаний. Наиболее известные открытые графы знаний: DBpedia [13], Google Knowledge Graph [14], YAGO [15]. Стандартные задачи, решаемые индустриальным графом знаний, перечислены ниже [16]:

- Построение цифровых двойников реального оборудования.
- Управление рисками.
- Мониторинг процессов.
- Операционные сервисы для сложного оборудования.

Для построения объединенной модели телекоммуникационной сети предлагается объединить структурные графовые модели сети в одном графе знаний. Это позволит устанавливать взаимосвязь динамических данных с данными статических сетевых моделей и между собой. В результате, обращаясь к графу знаний, будет возможно получить множество данных как о состоянии сети, так и о связях между различными ее элементами. При решении практических задач для телекоммуникационных сетей, предлагается рассматривать граф знаний как RDF (Resource Description Framework) граф, состоящий из RDF триплетов — «субъект — предикат — объект». В такой нотации множество RDF-утверждений образует ориентированный граф, в котором вершинами являются субъекты и объекты, а рёбра отображают отношения между ними [5][6]. Использование графов знаний может позволить устанавливать связи между традиционными графовыми моделями и создавать на их основе объединенную модель сети, позволяющую решать более сложные прикладные задачи. При различных сценариях одни и те же узлы могут являться как субъектами, так и объектами при их взаимодействии. Графы знаний позволяют в рамках одной модели связать, например, сервисную модель с моделью распространения прав доступа и моделью пользовательского интерфейса. Такой подход позволяет одним запросом к графу знаний получать ответ на такие вопросы как: по каким маршрутам пользовательского интерфейса выбранный пользователь сможет получить доступ к определенным данным и пр.

Существенным компонентом графа знаний телекоммуникационной сети являются динамические данные, в частности, данные о поведении пользователей.

VI. ПОТРЕБНОСТИ В ОБОБЩЕННЫХ МОДЕЛЯХ ПОЛЬЗОВАТЕЛЕЙ ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЕЙ

При определении потребностей пользователей в обобщенных моделях телекоммуникационной сети, рассматривалась сеть одного из крупных операторов кабельного ТВ. Обработка данных проводилась авторами в ходе выполнения проектов компанией Zodiac Systems [7] по заказам операторов кабельного ТВ Северной Америки. В результате анализа был получен перечень задач, для решения которых требуется построение обобщенных графовых моделей. Перечень в разрезе заинтересованных сторон представлен ниже:

1. Задачи конечных пользователей:
 - Получение актуальных данных об ограничениях для пользователя;
 - Поиск по доступным / всем данным;
2. Бизнес-задачи владельцев сети:
 - Получение информации об интересах пользователей;
 - Определение целевых групп пользователей для распространения рекламы;
3. Задачи служб эксплуатации сети:
 - Получение данных для быстрого решения инцидентов, возникающих у пользователей.

Использование графа знаний позволяет построить модель, отвечающую интересам всех потребителей данных о телекоммуникационных сетях. Такая модель включает: данные о конфигурации сети (на различных уровнях), данные о сетевых устройствах, пользователей, правах доступа, сетевых сервисах, клиентских приложениях и их версиях, статистику действий пользователей и другие.

VII. ПРОИЗВОДИТЕЛЬНОСТЬ СИСТЕМ НА БАЗЕ ГРАФОВ ЗНАНИЙ

Основываясь на данных исследований производительности систем на основе графа знаний [17], средняя скорость выполнения поисковых запросов к графу знаний при размере графа 1М триплетов, по которым производится поиск для наилучшего решения RDF хранилища (Virtuoso 7.2.4) не превышает 1сек., и может увеличиваться до 4 сек. под нагрузкой, при одновременном заполнении графа знаний новыми данными, что является приемлемым показателем для аналитических систем. При этом использование комбинированного решения, когда статические данные хранятся в RDF хранилище, динамические данные в SQL DBMS, а SPARQL запросы преобразуются в SQL для поиска динамических данных, не дает выигрыша.

VIII. ПРИМЕР ПОСТРОЕНИЯ ГРАФА ЗНАНИЙ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ

Одной из актуальных задач операторов телекоммуникационных сетей является адресная рассылка сообщений по группам пользователей. При определении групп учтены следующие условия:

- Проживание в определенном районе;
- Модель используемого абонентского устройства;
- Однократное или более частое использование сервиса PPV (Pay per view service – платный просмотр одной передачи) в течение выбранных суток для просмотра передач спортивной тематики.

На рис. 1 показана неформализованная диаграмма фрагмента данных в виде графа, достаточных для выполнения такого запроса. Диаграмма состоит из следующих узлов:

Статическая часть графа знаний:

- User_ID — идентификатор пользователя;
- Device_ID — идентификатор устройства;
- Hub_ID — идентификатор хаба, к которому присоединен Device_ID;
- Device_Model_ID — идентификатор модели;
- Service_ID — идентификатор вызываемого сервиса;
- Asset_ID — идентификатор информационного объекта, используемого сервисом;
- Asset_Genre_ID — идентификатор жанра.

Динамическая часть графа знаний:

- Request_ID — идентификатор запроса пользователя.

В качестве значений параметров для запроса выбраны:

Hub_ID="H000001";

Service_ID="PPV";

Asset_Genre_ID="Sport";

Дата: 08 февраля 2020г.

Для построения описанного графа знаний необходимо программное решение, удовлетворяющее следующим требованиям:

- наличие графовой базы данных с поддержкой SPARQL 1.1, что обеспечит реализацию более гибких запросов по сравнению с версией 1.0 [18];
- поддержка текстовых форматов JSON или XML для импорта данных, что обеспечит загрузку данных в графовую БД;
- наличие графического интерфейса для разработчика для анализа параметров графа знаний и скорости выполнения SPARQL запросов и загрузки данных;
- наличие возможности быстро получить поддержку со стороны разработчика.

Программное обеспечение должно быть свободно распространяемым или предоставлять соответствующую лицензию для проведения исследований.

Для построения графа знаний было выбрано решение Metaphactory (<https://www.metaphacts.com/product>), которое удовлетворяет всем перечисленным требованиям. Модель графа знаний включает:

- H000001—H000003 — идентификаторы хабов;
- Moto2k, Cisco3260, ArrisWB11, ArrisWB20 — идентификаторы моделей устройств;
- D000001—D000010 — идентификаторы устройств;

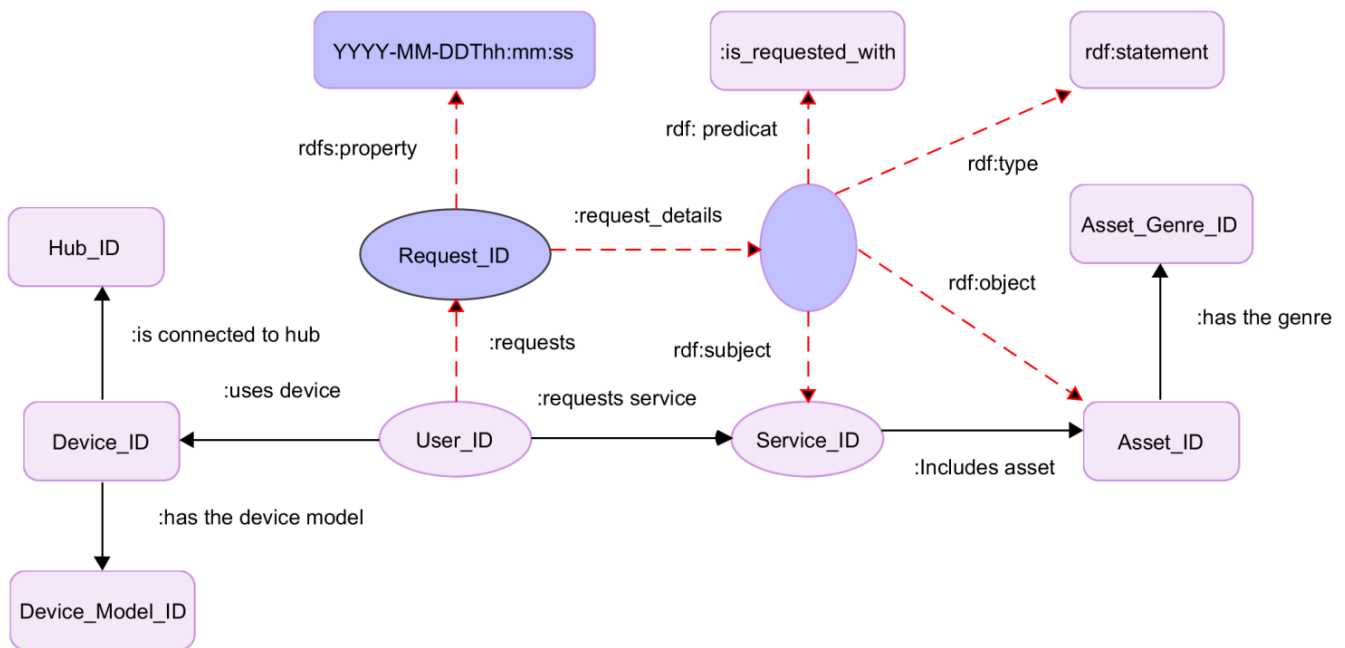


Рис. 1 Неформализованная диаграмма фрагмента описания графа знаний телекоммуникационной сети

- U000001—U000010 — идентификаторы пользователей;
- PPV, nDVR, CallerID — идентификаторы сервисов;
- Sport, Comedy, News — идентификаторы жанров ТВ программ;
- A000001—A000010 — идентификаторы ассетов (передач).

Граф знаний построен в соответствии с моделью, описанной выше. В ходе моделирования было подтверждено, что искомая группа пользователей может быть определена одним запросом.

Запрос к графу знаний имеет вид:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
PREFIX my: <http://127.0.0.1/bg/ont/test1#>
```

```
PREFIX : <http://www.metaphacts.com/resource/>
```

```
SELECT *
```

```
WHERE {
```

```
?Device my:is_connected_to_hub "H000001".
```

```
?Device my:has_the_device_model "Moto2k".
```

```
?Device my:has_id ?Device_id .
```

```
?User my:uses_device ?Device_id .
```

```
?User my:requests ?Request_ID .
```

```
?Request_ID rdf:property ?Date
```

```
FILTER contains(?Date, "2020-02-08") .
```

```
?Request_ID my:request_details ?Details_ID .
```

```
?Details_ID rdf:subject "PPV" .
```

```
?Details_ID rdf:object ?Asset_ID .
```

```
?Asset_ID my:has_the_genre "Sport"
```

```
}
```

Запрос к графу знаний возвращает список пользователей, которые используют устройства модели "Moto2k", присоединенные к хабу "H000001", и которые хотя бы один раз за сутки 08 февраля 2020г. приобретали разовый просмотр ТВ программы с жанром "Sport".

Выводы

Модели телекоммуникационных сетей, построенные на базе графов знаний, имеют практическую ценность в условиях, когда требуется объединить существующие графовые модели в рамках одной модели. За счет совместного использования нескольких графовых моделей, граф знаний позволяет решать более широкий класс задач. В докладе определены основные потребители данных о телекоммуникационных сетях, рассмотрены их интересы. Приведен пример построения графа знаний по данным телекоммуникационных сетей, показано решение практической задачи. Также представлена оценка производительности похожих систем на базе графа знаний.

Проведенный анализ возможностей и потребностей в применении графов знаний при построении моделей телекоммуникационных сетей показал, что такая задача является актуальной и имеет практическую значимость.

СПИСОК ЛИТЕРАТУРЫ

- [1] Spatial and Graph Topology Data Model and Network Data Model Graph Developer's Guide
<https://docs.oracle.com/database/121/TOPOL/network-data-model-graph-overview.htm#TOPOL700>
- [2] Sudip Saha, Mahantesh Halappanavar, Anil Vullikanti. Identifying Vulnerabilities and Hardening Attack Graphs for Networked Systems, Virginia Tech, 2014
http://staff.vbi.vt.edu/ssaha/papers/attackgraph_dag.pdf
- [3] Lipton, Richard J.; Snyder, Lawrence (1977). "A Linear Time Algorithm for Deciding Subject Security" (PDF). *Journal of the ACM*. 24 (3): 455–464. CiteSeerX 10.1.1.149.4807. doi:10.1145/322017.322025
- [4] M. Kroetsch and G. Weikum. *Journal of Web Semantics: Special Issue on relationships between entities.* Knowledge Graphs. <http://www.websemanticsjournal.org/index.php/ps/announcement/view/19> [August, 2016]
- [5] RDF Primer <https://www.w3.org/TR/rdf-primer/>
- [6] M. Farber, B. Ell, C. Menne, A. Rettinger, and F. Bartscherer. Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web Journal*, 2016. <http://www.semantic-web-journal.net/content/linking-data-quality-dbpedia-freebase-opencyc-wikidata-and-yago> [August, 2016] (revised version, under review)
- [7] Zodiac Systems Inc web-site: <https://www.zodiacsystems.com>
- [8] Splunk: <https://www.splunk.com/>
- [9] Datadog: <https://www.datadoghq.com/>
- [10] Logstash: <https://www.elastic.co/logstash>
- [11] IBM Telecom Analytics Solutions: <https://www.ibm.com/support/pages/ibm-telecom-analytics-solutions-product-documentation>
- [12] Huawei Big Data Analytics (BDA): <https://carrier.huawei.com/en/products/carrier-software/enabling-digital-operation-business-agility/fusioninsight-universe-big-data-analytics>
- [13] DBpedia: <http://wiki.dbpedia.org/>
- [14] Google knowledge graph: <http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>
- [15] YAGO: <http://mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/>
- [16] Hubauer T. et al. Use Cases of the Industrial Knowledge Graph at Siemens. *International Semantic Web Conference (P&D/Industry/BlueSky)*, 2018
- [17] Pierfrancesco Bellini, Paolo Nesi, "Performance assessment of RDF graph databases for smart city services". DOI: <https://doi.org/10.1016/j.jvlc.2018.03.002>

[18] <https://www.w3.org/TR/sparql-features/>

Practical aspects of using knowledge graphs for telecommunication networks modelling

Nataly Zhukova

Dept. of Software Engineering & Computer Applications (MOEVM)

St. Petersburg State Electrotechnical University, St. Petersburg, Russia,

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russia,

nazhukova@mail.ru

ORCID: 0000-0001-5877-4461

Igor Kulikov

Dept. of Software Engineering & Computer Applications (MOEVM)

St. Petersburg State Electrotechnical University, St. Petersburg, Russia,

i.a.kulikov@gmail.com

ORCID: 0000-0002-2532-5579

Abstract: The paper deals with the principles and practices of using graphs as a tool for modelling telecommunication networks. It contains a review of network graph models currently used, discusses network characteristics which can be analyzed using these models, and shows the advantages and disadvantages of these models. It also contains an analysis of present-day requirements that telecommunication network models should meet, such as the opportunity to exercise dynamic control over network parameters, access to information, and the quality of services provided by networks to their end users. It has been assessed whether the graph models being used meet the current and potential requirements. The paper gives a definition of a knowledge graph and assesses the prospects for using models based on knowledge graphs in modelling telecommunication networks. It identifies problems which can be solved using models based on knowledge graphs.