О ФОРМАЛИЗАЦИИ ОПИСАНИЯ КОМПЬЮТЕРНОЙ СЕТИ ДЛЯ ДИАГНОСТИКИ НА ОСНОВЕ МЕТОДОВ РАСПОЗНАВАНИЯ

Е.В. Олизарович, В.Г. Родченко

Гродненский государственный университет им. Я.Купалы,
кафедра программного обеспечения интеллектуальных и компьютерных систем
230023 ул. Ожешко,22, г. Гродно, Республика Беларусь
телефон: + 375 152 731907; факс: + 375 152 731910; e-mail: e.olizarovich@grsu.by
web: www.grsu.by

Рассматривается метод формализации описания компьютерной сети для применения в задачах классификации состояний. В основе построения математической модели компьютерной сети лежит двухступенчатое преобразование результатов первичных измерений трафика. Полученное формальное описание ориентировано на обработку методами математической теории распознавания образов.

Ключевые слова – диагностика, компьютерная сеть, моделирование, распознавание образов

1 ВВЕДЕНИЕ

Существующие средства и методы технической диагностики компьютерных сетей (КС) предлагают широкий спектр возможностей по измерению параметров и определению их отклонений от нормативных значений. Однако в специальной литературе и на рынке диагностического оборудования мало представлены технологии и инструменты динамического анализа интегральных параметров функционирования КС, как сложной системы, взаимодействующей с другими объектами.

Наиболее распространенные подходы к организации диагностики КС, как правило, ориентированы на измерение и оценку стандартных технических показателей, заложенных при разработке оборудования. Показатели функционирования включают характеристики среды и канала передачи, суммирующие и статистические параметры и т.д. Для диагностики трафика КС наиболее часто применяются методы, основанные на использовании протокола управления SNMP. Существующие базы данных значений, обрабатываемых протоколом SNMP, содержат тысячи элементов и постоянно возрастают. Такое количество показателей как правило избыточно для решения конкретных задач, что увеличивает трудоемкость операций и снижает общую эффективность затрат на управление системой. При этом, основная доля аналитической работы, как правило, возлагается на человека, выполняющего обслуживание сети, а возможности измерительных средств ограничиваются определенной спецификацией сетевого оборудования или заданной сетевой технологией.

Методы и средства управления сетевыми ресурсами быстро устаревают, т.к. постоянно возникают качественно новые диагностические задачи, вызванные такими тенденциями развития, как расширение области применения КС, усложнение структуры, расширение номенклатуры и смена поколений программно-технических технологий взаимодействия.

При диагностике состояний и событий, относящихся к уровням управления сетью и услугами, возникают следующие проблемы построения моделей КС:

- требуется анализ объемных массивов входных параметров и классификация большого числа теоретически вероятных состояний КС;
- в течение срока жизни требуется постоянное изменение системы диагностики в связи с характерными для вычислительной техники процессами постоянного развития структуры, модернизации программного обеспечения, и совершенствования сервисных функций;
- существующие расчетные модели КС не позволяют строить пригодные для практической диагностики, описания сложных сетей, поскольку требуют недоступной или неизвестной информации;
- существующие технологии диагностики не предоставляют возможностей адаптации на основе апостериорной (эксплуатационной) информации о КС.

Для решения подобных задач в разных областях науки и техники с успехом применяются методы математической теории распознавания образов [1,2]. Одной из проблем применения таких методов является поиск информативных признаков. В ряде случаев в качестве признаков распознавания выбираются не измеряемые внешние признаки систем, а синтезированные на их основе сложные характеристики. Наиболее известным примером могут служить томографы и радиотелескопы.

Настоящий доклад посвящен описанию метода формализации состояния компьютерной сети, основанному на анализе апостериорной информации о трафике КС. Обосновывается адекватность предложенной математической модели (ММ) КС, приводятся примеры правил и алгоритмов расчета ее параметров.

Предлагаемый метод формализации ориентирован на применение в автоматизированных системах диагности-

ки, базирующихся на основе методов математической теории распознавания образов [3].

2 МЕТОДЫ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Как одна из подзадач управления, диагностика оперирует не реальными объектами, а их математическими моделями. При этом точность и эффективность методов определяется адекватностью применяемых моделей.

В общем виде состояние компьютерной сети $S_{КС}$, может быть представлено в виде

$$\vec{Y}_{КС} = f_{КС}(\vec{X}_{КС}, \vec{V}_{КС}, \vec{H}_{КС}). \quad (1.1)$$

где $X_{КС}=\{x_1\ldots,x_{k1}\}$- множество входных воздействий, $V_{КС}=\{v_1,\ldots,v_{k2}\}$ – множество воздействий внешней среды, $H_{КС}=\{h_1,\ldots,h_{k3}\}$ - совокупность собственных (внутренних) параметров; $Y_{КС}=\{y_1,\ldots,y_{k4}\}$ - совокупность выходных характеристик КС, $f_{КС}$ - функция, логические условия, алгоритм, таблица или описание правил.

Аналитическое представление состояния действующей КС в виде (1.1), как правило, невозможен, поскольку необходимо учитывать неизвестные или неизмеримые параметры. Методы обучения теории распознавания образов позволяют стоить эмпирические зависимости $f_{КС}$ и на их основе предсказывать поведение сети.

Метод диагностики на основе анализа трафика предполагает классификацию состояния КС на основе информации о внешних измеримых характеристиках. Задачей формализации является построение такого множества $Y_{ММКС}$, элементы которого будут составлять рабочий словарь признаков системы распознавания, а для вектора состояния выполняется

$$\vec{Y}_{ММКС} \equiv f_{КС}(\vec{X}_{КС}, \vec{V}_{КС}, \vec{H}_{КС}). \quad (1.2)$$

Как правило в процессе диагностики требуется определить состояние вектора $\vec{H}_{КС}$, а граница "система — среда" может быть выбрана так, чтобы вид элементов множеств $X_{КС}$ и $V_{КС}$ был априорно известен. Тогда, возможно построение адекватной модели диагностики КС как эмпирической зависимости от измеримых выходных параметров КС:

$$\vec{S}_{КС} = f'_{КС}(\vec{Y}_{ММ\,КС}). \quad (1.3)$$

Задачей диагностики является обнаружение закономерностей изменения $\vec{Y}_{ММКС}$ при изменении состояния КС и построение комплекта эталонных наборов значений $\vec{Y}_{ММКС\,i}$ для искомых состояний s_i.

Важным достоинством применения методов распознавания, как математического инструмента, является возможность отделения физической природы характеристик КС от их информативной значимости. При этом, система распознавания оперирует алфавитом признаков – множеством вторичных параметров $Y_{ММКС}$, полученных путем предварительной математической обработки множества $X_{ММКС}$ первичных, т.е. инструментально измеренных характеристик. Применение двухуровневой схемы подготовки данных позволяет достичь следующих целей:

- метод диагностики не зависит от средств измерений;
- возможно получение новых знаний о свойствах объекта, на основе параметров, которые недоступны для методов прямых измерений.

При диагностике состояния КС количество возможных значений может быть неопределенно большим, поэтому, как правило, необходимо выделить подмножества состояний и соответствующих выходных характеристик, которые необходимы и достаточны для решения конкретной диагностической задачи. При этом, отброшенные параметры не должны существенно влиять на результат диагностики.

Как следует из формулы (1.3), основная задача формализации КС – построение ММКС, адекватно отражающей выходные параметры КС в соответствии с условиями диагностической задачи. В обобщенном виде процесс построения ММКС может быть представлен в виде

$$X_{ММКС} \xrightarrow{V_{Ммкс}, H_{ММКС}} Y_{ММКС}. \quad (1.4)$$

где $X_{ММКС}$ – множество измеримых параметров – набор технических и информационных характеристик трафика КС, значения которых могут быть измерены программно-техническими средствами; $V_{ММКС}$ – алфавит состояний, которые должны распознаваться, $V_{ММКС} \subset S_{КС}$; $H_{ММКС}$ – множество методов (алгоритмов) преобразований измеряемых параметров $X_{ММКС}$ в словарь признаков $Y_{ММКС}$. В процессе итерационного построения ММКС множества $X_{ММКС}$, $Y_{ММКС}$, $H_{ММКС}$, $V_{ММКС}$ могут уточняться.

Поскольку точность и гибкость метода диагностики возрастает с увеличением словаря, то на первом этапе построения ММКС основной задачей является нахождение наибольшего количества $X_{ММКС}$ и $Y_{ММКС}$ - характеристик КС, логически связанных с диагностируемыми состояниями. Однако, при этом значительно возрастает объем вычислений, поэтому необходимо обеспечить вхождение в рабочий алфавит только наиболее информативных признаков. В общем случае, процесс построения ММКС является двухфазным: построение априорного словаря признаков $Y'_{ММКС}$ и расчет рабочего словаря $Y_{ММКС}$, где $Y_{ММКС} \subset Y'_{ММКС}$. Соответственно в процессе построения ММКС условно можно выделить два набора данных: априорный $X'_{ММКС}$, $Y'_{ММКС}$, $H'_{ММКС}$ и рабочий $X_{ММКС}$, $Y_{ММКС}$, $H_{ММКС}$, где $X_{ММКС} \subset X'_{ММКС}$, $Y_{ММКС} \subset Y'_{ММКС}$, $H_{ММКС} \subset H'_{ММКС}$.

Априорные множества $X'_{ММКС}$ и $Y'_{ММКС}$ строятся на основе знаний об источниках первичных данных и в зависимости от имеющихся возможностей измерения трафика. Методы формирования:

1. Типовой - элементы множеств $X'_{ММКС}$ и $Y'_{ММКС}$ выбираются из типовых библиотек, созданных на основе опыта решения подобных диагностических задач;

2. Экспертный - априорные множества формируются на основе предположений экспертов.

Задачей второго этапа построения ММКС является сокращение априорного словаря признаков $Y'_{ММКС}$ и формирование его подмножества $Y_{ММКС}$, путем исключения неклассифицирующих и зависимых параметров. Для этого, на основе априорных экспертных заключений или по результатам анализа контрольной выборки строится квалифицированная обучающая выборка и формируется рабочий словарь признаков [4].

Для построения $Y_{ММКС}$ и $H_{ММКС}$ требуется анализ совокупностей векторов $\vec{X}_{ММКС}(t)$, наблюдаемых в течение заданного периода Δt_x . Величина Δt_x, зависит от особенностей диагностической задачи и соответствует минимальному периоду наблюдения, в течение которого КС должна проявить все значимые для задачи диагностики свойства. Для компьютерной сети, находящейся в стационарном состоянии, результат диагностики для каждого периода Δt_x должен быть одинаков.

Типовыми элементами множества преобразований $H_{ММКС}$ являются следующие операции:

1. Агрегация и фильтрация:

– суммирование численных значений $x_{ммкci}$ за период Δt_x:

$$h(\Delta t) = \sum x_{ММКСi} \,. \tag{1.5}$$

– подсчет количества событий $x_{ммкci} = A$, зафиксированных средствами измерения за период Δt_x, $i \in \{1,2,\ldots,|X_{ММКС}|\}$:

$$h(\Delta t) = \sum r \text{, где} \begin{cases} r = 1, & при \;\; x_{ММКСi} = A; \\ r = 0, & при \;\; x_{ММКСi} \neq A; \end{cases} \tag{1.6}$$

– подсчет количества зафиксированных за период Δt_x событий $x_{ммкci} = A_1$, при условии выполнения $x_{ммкcj} = A_2$, $i, j \in \{1,\ldots, |X_{ММКС}|\}$, $j \neq i$:

$$h(\Delta t) = \sum r, \begin{cases} r = 1, при \;\; x_{ммкci} = A_1 \wedge x_{ммкcj} = A_2; \\ r = 0, при \;\; x_{ммкci} \neq A_1 \vee x_{ммкcj} \neq A_2; \end{cases} \tag{1.7}$$

– приведение к эквидистантной шкале, т.е. вычисление частоты (плотности) событий в единицу времени Δt:

$$p = h(\Delta t) / \Delta t \tag{1.8}$$

2. Расчет статистических характеристик выборки:

– выборочное среднее значение

$$\bar{x}_{ММКС} = \frac{1}{n} \sum_{i=1}^{n} x_{ММКСi} \tag{1.9}$$

где n – объем выборки измерений за период Δt_x;

– выборочная дисперсия

$$h = \overline{x^2_{ММКС}} - (\bar{x}_{ММКС})^2 \tag{1.10}$$

где n – объем выборки измерений за период Δt_x.

3. Нормирование:

$$y^*_{ММКС_i} = \frac{y_{ММКС_i} - \min\limits_{i=1,n}(y_{ММКС_i})}{\max\limits_{i=1,n}(y_{ММКС_i}) - \min\limits_{i=1,n}(y_{ММКС_i})} \tag{1.11}$$

Математическая модель КС, разработанная для целей диагностики на основе методов математической теории распознавания образов включает:

- $X_{ММКС}$ - множество первичных параметров;
- $Y_{ММКС}$ - пространство признаков состояния КС, которое будет использовано для расчетов;
- $H_{ММКС}$ – множество математических способов формирования пространства признаков.

3 ФОРМАЛИЗАЦИЯ ОПИСАНИЯ КОМПЬЮТЕРНОЙ СЕТИ

Процесс передачи данных - многоуровневый процесс. Согласно принятой эталонной модели взаимодействия открытых систем, каждая единица информации, передаваемая в КС, может быть условно представлена в виде иерархии заголовков. Измеримым результатом воздействия каждого уровня на передаваемый пакет данных является содержание соответствующего заголовка, а каждый кадр, передаваемый в канале, содержит информацию обо всех сеансах, задействованных на разных уровнях в процессе взаимодействия клиента с сервером. Первичными источниками данных о состоянии и процессах, происходящих в исследуемом сегменте КС, могут служить как собственно значения полей в заголовках, так и их статистические характеристики.

Структура кадров позволяет построить обобщенное формальное описание трафика компьютерной сети на основе многоуровневой иерархии измеримых параметров

$$X_{ММКС} = X_1 \cup X_2 \cup X_3 \cup X_4 \cup X_5 \cup X_6 \cup X_7 \tag{1.12}$$

где $X_i = \{x_{i1},\ldots,x_{in_i}\}$, наборы полей заголовков i-го уровня взаимодействия, n_i - количество полей, предусмотренных для заголовка i-го уровня. Множество $X_{ММКС}$ составляют $n = \sum_{i=1}^{7} n_i$ детерминированных характеристик, определяемых стандартами и спецификациями соответствующих протоколов. В предельном случае, множество $X_{ММКС}$ представляет собой универсальное множество $X^0_{ММКС}$ – генеральную совокупность параметров, включающую все возможные поля заголовков.

Источником получения данных о техническом состоянии и информационной структуре компьютерной сети являются кадры (frame), передаваемые устройствами MAC-уровня. В качестве измеряемых параметров могут

выступать значения полей заголовков пакетов различных уровней взаимодействия, содержащиеся в кадрах. Наибольшую информативность при испытаниях проявили модели, базирующиеся на следующих уровнях: канальный (Data Link), сетевой (Network), транспортный (Transport), прикладной (Application). При анализе различных аспектов состояния сети в качестве элементов множества $X_{ММКС}$ могут быть использованы поля: тип протокола передачи сетевого уровня, тип протокола передачи транспортного уровня, MAC - адрес источника кадра; MAC - адрес получателя; формат кадра; IP - адрес источника пакета; IP - адрес получателя; номер порта; длина кадра [5]. Технически процесс наблюдения реализуется путем создания и анализа первичных источников информации – журналов работы различных измерительных систем. Все данные должны быть извлечены из соответствующих полей заголовков канального, сетевого и транспортного уровней OSI и занесены в таблицу журнала с указанием времени регистрации кадра.

В качестве элементов множества $Y_{ММКС}$ могут быть использованы такие параметры как частота передачи кадров, частота передачи IP-пакетов, частота появления ARP запросов, суммарное количество исходящих пакетов от определенного адреса, количество различных IP-адресов-источников пакетах исходящих от одного MAC-адреса, суммарный объем кадров исходящих от определенного адреса, средняя длина кадров, частота появления широковещательных кадров.

4 ЗАКЛЮЧЕНИЕ

В современном понимании диагностика КС, как задача управления, должна быть направлена не только на обнаружение неисправностей отдельных узлов, а также на оценку прикладных свойств и повышение эффективности функционирования сети в целом. Предложенный метод формализации предназначен для решения задач диагностики в области управления сетью и услугами. Данный круг задач является сегодня наименее освоенным, но его актуальность возрастает с усложнением сетей и переходом от понятия телекоммуникационной сети к сети информационно-технической.

Преимуществом метода является более высокая степень универсальности в сравнении с существующими, т.к. он ориентирован на применение в задачах, для которых отсутствуют типовые решения: сети с неуправляемым оборудованием, гетерогенные сети, контроль работы пользователей, обнаружение скрытых процессов. В отличие от большинства существующих подходов, разработанный метод диагностики основан на структурных принципах организации трафика и может быть применен в любой сетевой системе соответствующей архитектуры. Таким образом, метод диагностики работоспособен в любой низкобюджетной или устаревшей КС, поскольку не зависит от фирменных технических решений, модели оборудования, специальных протоколов передачи, но, при необходимости может дополнительно использовать их функциональность.

Предлагаемый метод формализации ориентирован на использование в составе автоматизированных систем диагностики, основанных на использовании аппарата математической теории распознавания образов, но не зависит от них, поэтому может применяться и в других случаях.

Достоинством метода является возможность построения "открытых" систем диагностики, т.е. систем позволяющих создавать и изменять библиотеки признаков, алгоритмов преобразования. К преимуществам также относится возможность диагностики без оказания воздействия на сеть и возможность "отложенной" диагностики на основании анализа журналов работы.

В качестве ограничения предложенного метода формализации следует отметить зависимость эффективности от качества экспертного заключения на всех этапах разработки модели.

Предложенные решения могут быть широко использованы, поскольку не требуют специального дорогостоящего измерительного оборудования, а базируются на стандартных платформах и интерфейсах вычислительных машин.

ЛИТЕРАТУРА

[1] Загоруйко, Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. – Новосибирск: Изд-во Института математики СО РАН, 1999. – 268 с.

[2] Симанков, В.С. Адаптивное управление сложными системами на основе теории распознавания образов: Монография (научное издание) / В.С. Симанков, Е.В. Луценко. – Техн. ун-т Кубан. гос. технол. ун-та. – Краснодар, 1999. –318 с.

[3] Жукевич, А.И. Метод построения эталонов состояний компьютерной сети на основе применения алгоритмов теории распознавания образов /А.И. Жукевич, Е.В. Олизарович, В.Г. Родченко// Сетевые компьютерные технологии: сб.тр. III Междунар.науч.конф., 17-19 окт.2007г. Минск /редкол.: М.К.Буза (отв.ред), А.Н.Курбацкий [и др.]. – Минск: Изд.центр БГУ, 2007. – С.14-17.

[4] Олизарович, Е.В. Построение концептуальной модели диагностики технической системы по результатам наблюдений на основе методов математической теории распознавания образов /Е.В. Олизарович// Известия Гомельского государственного университета имени Ф.Скорины. – 2006. – №4(37). – С. 58-61.

[5] Олизарович, Е.В. О применении методов распознавания в задачах практической диагностики режимов работы узлов компьютерной сети /Е.В.Олизарович// Известия Гомельского государственного университета имени Ф.Скорины. – 2007. – №5(44). - С.53-57.