УДК 681.003.66

С.А. СОКОЛОВ, А.Л. СТОКИПНЫЙ, А.В. ГОЛДАЕВ

ДИАГНОСТИЧЕСКАЯ МОДЕЛЬ IP-СЕТИ

Дается теоретическое описание диагностической модели IP-сети, которая позволяет представить состояние сети и ее служб в виде набора событий. Каждое событие рассматривается как факт принятия некоторым диагностическим параметром значения из ранее заданного подмножества значений области определения. Описание состояния сети формируется в детерминированные моменты времени. Вводится понятие базы данных диагностической информации, которая состоит из совокупности описаний состояний IP-сети на некотором временном интервале. Проводится классификация источников диагностической информации и определяются типы диагностических параметров.

1. Введение

Современные высокоэффективные системы сбора, обработки, передачи и хранения информации в подавляющем большинстве случаев имеют распределенный характер и используют для организации передачи информации пакетные сети, построенные на базе IP-технологий. Ввиду этого важность мероприятий, связанных с контролем функционирования, мониторингом состояния отдельных устройств IP-сети, своевременным реагированием на факты возникновения неисправностей с последующими процедурами локализации и восстановлении, является обоснованной.

Неисправное состояние сети передачи данных ведет к невозможности выполнять заданные функции сетевыми службами – основными структурными элементами современных информационных систем. Здесь под службой авторы понимают прикладную задачу, компоненты которой могут выполняться в отдельных аппаратно-программных окружениях и взаимодействуют посредством сети передачи данных. Одним из основных показателей информационной системы и, следовательно, IP-сети как ее составной части, является «доступность» или «коэффициент готовности» [1,2]:

$$K_{\Gamma} = \frac{T_{\Sigma}}{T_{\Sigma} + T_{НПЛ}}, \qquad (1)$$

где T_{Σ} – суммарное время, в течение которого сеть работоспособна; $T_{П}$ – суммарное время простоя.

Исходя из порядка расчета K_{Γ} (1), повышение «доступности» означает минимизацию значения $T_{П}$. Одним из способов уменьшения значения суммарного времени простоя является внедрение современных диагностических средств и эффективных методик диагностирования.

Существующие на сегодня методы диагностирования IP-сети включают в себя многоадресную и одноадресную томографию [3,4,5], выявления неисправностей сети на основе вероятностных рассуждений [6], применение экспертных систем [7] и нейросетевые решения [8]. Перспективным направлением является применение методов Data Mining для поиска закономерностей в работе программных и аппаратных компонентов IP-сетей [9].

Цель исследования – разработка диагностической модели IP-сети, которая бы позволяла описать состояние как всей сети, так и отдельной службы в виде, пригодном для формирования на заданном временном интервале базы данных диагностической информации (БДДИ) IP-сети. Полученная БДДИ в дальнейшем может быть использована для выявления с использованием методологии Data Mining закономерностей между значениями диагностических параметров и состоянием служб IP-сети.

2. Разработка ситуационной модели IP-сети

В основу разрабатываемой модели положим представление IP-сети в виде множества служб $S = \{s_1, s_2 \ldots s_l \ldots s_L\}, l = \overline{1, L}$, где L – общее количество служб.

Определение 1. Источником диагностической информации (ИДИ) называется компонент IP-сети, предоставляющий необходимую для определения состояния службы IP-сети информацию в виде диагностических параметров.

Выделим следующие классы ИДИ, которые отличаются по характеру предоставляемой диагностической информации:

– Программные компоненты. В системе диагностики представлены через параметры, которые описывают динамику функционирования вычислительного процесса.

– Аппаратные компоненты. Рассматриваются как множество параметров, отражающих состояние таких объектов как процессор, память, сетевой интерфейс, устройство хранения данных.

– Каналы связи. В качестве диагностических параметров выступают характеристики физической или логической (в случае использования виртуальных каналов) среды передачи между двумя сетевыми интерфейсами. Для получения характеристик обычно используется сетевой анализатор, подключенный непосредственно к каналу связи, или программные средства на основе протокола ICMP.

Далее в работе в качестве ИДИ мы будем рассматривать только те компоненты, которые предоставляют доступ к значениям диагностических параметров посредством протокола SNMP. Данное условие не сужает область применения модели, поскольку на сегодняшний день подавляющее большинство аппаратных и программных компонентов содержит реализацию указанного протокола.

Обозначим совокупность ИДИ в виде множества $A := \{a_1, a_i, ...a_I\}, i = \overline{1, I}$, где I – общее количество ИДИ в рассматриваемой IP-сети. ИДИ, поддерживая одну или несколько MIB, может реализовать описанные в ней SNMP-объекты управления в виде переменных, которые содержат текущие значения соответствующих параметров. В данной модели SNMP-объекту управления, значения которого доступны для контроля, может быть поставлен в соответствие один или более диагностических параметров.

Рассмотрим SNMP-объект «Число пакетов, полученных с ошибкой», который является обязательным в первой и второй версиях MIB, имеющих на сегодняшний день статус стандартов Интернета (RFC 1156 и RFC 1213). В RFC 1213 переменная, которая должна реализовать указанный выше объект управления, имеет тип «counter» и содержит количество пакетов, полученных с ошибками на текущий момент времени. Упомянутому SNMP-объекту могут быть сопоставлены два диагностических параметра: первый – «число пакетов, полученных с ошибкой», второй – «скорость получения пакетов с ошибками». Последний представляет собой плотность или скорость появления ошибок на определенном промежутке времени и, с точки зрения диагностирования, содержит более полезную информацию.

Диагностические параметры (ДП) по характеру описания свойств компонентов и каналов передачи данных IP-сети можно классифицировать на общие и частные.

Определение 2. Частным диагностическим параметром (ЧДП) называется параметр, который отражает индивидуальную характеристику отдельно взятого аппаратного или программного компонента.

Примерами ЧДП являются процент загрузки процессора, объем свободной оперативной памяти, количество пакетов, принятых с ошибками, количество открытых TCP соединений.

Определение 3. Общим диагностическим параметром (ОДП) называется параметр, который содержит агрегированную оценку функционирования системы взаимодействующих компонентов.

К классу ОДП относятся такие параметры, как время задержки доставки пакета, доступность службы, джиттер, время реакции при обращении к службе, скорость передачи данных.

Контролируемые диагностические параметры a_i представим в виде множества:

$$\forall a_i \in A \; \exists B_i = \{b_{i1}, b_{i2}, \ldots, b_{ij_i}, \ldots, b_{iJ_i}\}, \quad (2)$$

где $j_i = \overline{1, J_i}$, J_i – количество диагностических параметров, предоставляемых ИДИ a_i.

Общее количество контролируемых параметров всех ИДИ:

$$N^B = \sum_{i=1}^{I} J_i . \qquad (3)$$

На основе данных технической документации и результатов наблюдения за процессом функционирования IP-сети человек-эксперт может сформировать множество диагностических параметров $B(s_l)$, значения которых в дальнейшем необходимо учитывать в процессе идентификации состояния некоторой службы s_l. Если для всех служб известны множества $B(s_l)$, то для любого параметра b_{ij_i} можно определить множество $S(b_{ij_i}) \in S$. Служба s_l принадлежит множеству $S(b_{ij_i})$ только при условии, что параметр b_{ij_i} используется в процессе определения ее состояния. Следует заметить, что правильность формирования $B(s_l)$ и $S(b_{ij_i})$ зависит, в первую очередь, от профессионального опыта, умения и навыков человека-эксперта.

Определение 4. Собственным параметром службы s_l называется параметр b_{ij_i}, для которого $\left|S(b_{ij_i})\right| = 1$ и $s_l \in S(b_{ij_i})$.

Определение 5. Неопределенным параметром службы s_l называется параметр b_{ij_i}, для которого $s_l \notin S(b_{ij_i})$ или же $s_l \in S(b_{ij_i})$ и $\left|S(b_{ij_i})\right| > 1$.

«Менеджер» посредством SNMP периодически обращается к «агенту», работающему в аппаратно-программном окружении ИДИ a_i, с запросом на получение текущего значения параметра b_{ij_i}. Теоретически, полученное от «агента» значение запрашиваемого параметра b_{ij_i} может быть использовано для описания некоторого состояния службы $s_l \in S(b_{ij_i})$. Однако, учитывая, что тип переменной, которая реализует SNMP-объект управления, в большинстве случаев является целочисленным на интервале $[1\ldots4294967295]$, допустимое количество состояний службы s_l составляет $4294967295^{|B(s_l)|}$. Идентифицировать такое количество состояний для сохранения и проведения последующего анализа является трудоемкой процедурой и нецелесообразно. На практике человек-эксперт разбивает все множество значений параметра на конечное число подмножеств. Каждое подмножество включает, по его мнению, значения, которые параметр принимает при одном и том же состояния служб $s_l \in S(b_{ij_i})$.

Пусть $D(b_{ij_i})$ – множество значений параметра b_{ij_i}, тогда его разбиение может быть записано в виде:

$$D(b_{ij_i}) = \bigcup_{k_{ij_i}}^{K_{ij_i}} D_{ij_i k_{ij_i}}(b_{ij_i}) , \qquad (4)$$

где K_{ij_i} – общее количество подмножеств; $D_{ij_i k_{ij_i}}(b_{ij_i})$ – подмножество значений параметра b_{ij_i}. Чем больше K_{ij_i}, тем более точно будет отслеживаться динамика функционирования служб $s_l \in S(b_{ij_i})$.

В данной модели (рис.1) примем, что множество $D(b_{ij_i})$, а следовательно, и подмножества $D_{ij_i k_{ij_i}}(b_{ij_i})$, являются вполне упорядоченным на основе отношения $<$ строгого порядка. Мощность и состав $D(b_{ij_i})$ зависят от типа, указанного в описании SNMP-объекта управления, которому соответствует параметр b_{ij_i}.

Рис. 1. Пример задания функций $c_{ij_ik_{ij_i}}(b_{ij_i})$, $K_{ij_i} = 3$, на множестве значений $D(b_{ij_i})$ параметра b_{ij_i} ИДИ a_i

Для каждого b_{ij_i} определим множество характеристических функций C_{ij_i}, принимающих в качестве аргумента значение из множества $D(b_{ij_i})$:

$$\forall b_{ij_i} \in B_i \ \exists C_{ij_i} = \{c_{ij_i1}(b_{ij_i}), c_{ij_i2}(b_{ij_i}), \ldots, c_{ij_ik_{ij_i}}(b_{ij_i}), \ldots, c_{ij_iK_{ij_i}}(b_{ij_i})\}, \quad (5)$$

где $k_{ij_i} = \overline{1, K_{ij_i}}$, K_{ij_i} – количество функций, определенных на множестве $D(b_{ij_i})$.

Характеристическая функция обладает следующими свойствами:

– Область значений представляет собой множество $\{0,1\}$.

– Являясь сурьективной, функция $c_{ij_ik_{ij_i}}(b_{ij_i})$ принимает единичные значения на множестве $D_{ij_ik_{ij_i}}(b_{ij_i})$, а нулевые – на множестве $D(b_{ij_i}) / D_{ij_ik_{ij_i}}(b_{ij_i})$.

– В общем случае имеет вид:

$$c_{ij_ik_{ij_i}}(b_{ij_i}) = \begin{cases} 1, \text{значение } b_{ij_i} \text{ принадлежит множеству } D_{ij_ik_{ij_i}}(b_{ij_i}); \\ 0, \text{ значение } b_{ij_i} \text{ принадлежит множеству } D(b_{ij_i}) / D_{ij_ik_{ij_i}}(b_{ij_i}). \end{cases} \quad (6)$$

Для каждого b_{ij_i} определим вектор, который содержит результаты выполнения функций из C_{ij_i}:

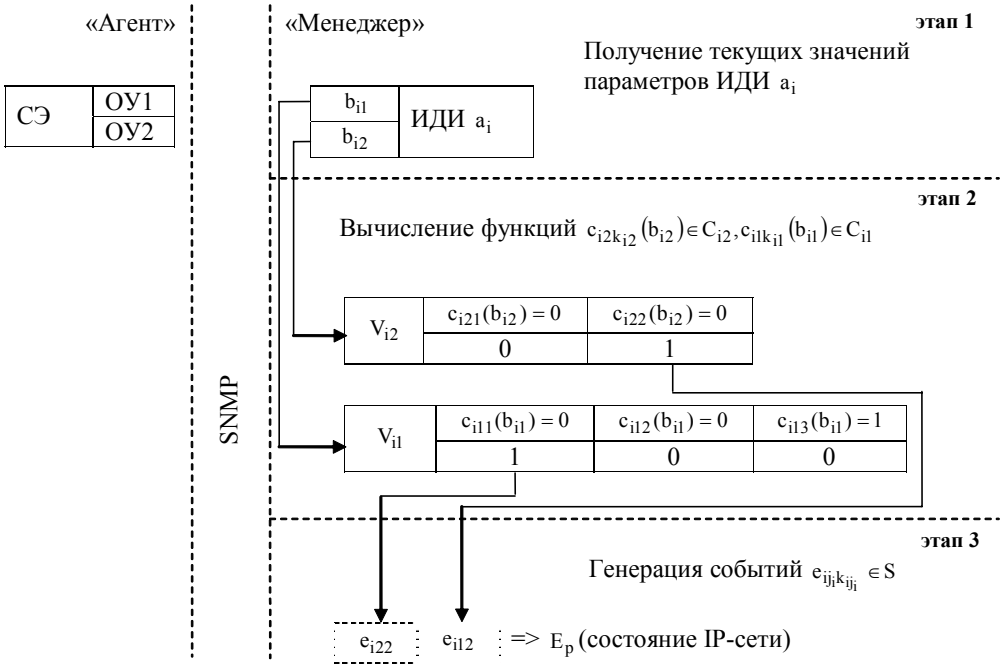
$$\forall b_{ij_i} \in B_i \ \exists V_{ij_i} = [v_{ij_i1} \in \{0,1\}, v_{ij_i2} \in \{0,1\}, \ldots, v_{ij_ik_{ij_i}} \in \{0,1\}, \ldots, v_{ij_iK_{ij_i}} \in \{0,1\}]. \quad (7)$$

Определение 6. Факт получения вектором V_{ij_i} результатов выполнения функций из C_{ij_i} называется событием.

Исходя из указанных ранее свойств функции $c_{ij_ik_{ij_i}}(b_{ij_i})$, справедливо следующее:

$$\forall V_{ij_i} \sum_{k=1}^{K_{ij_i}} v_{ij_ik} = 1, \quad (8)$$

т.е. в заданный момент времени для объекта a_i в отношении параметра b_{ij_i} только одна функция из C_{ij_i} принимает единичное значение. Следовательно, общее количество событий, инициируемых в отношении заданного параметра, равно K_{ij_i}.



СЭ – структурный элемент ТС
ОУ – SNMP-объект управления

Рис. 2. Модель генерации события $e_{ij_ik_{ij_i}} \in E$

Из сказанного выше следует, что для обозначения индекса события, которое инициировано параметром b_{ij_i} объекта a_i, можно использовать систему индексирования «объект – параметр – функция». Множество событий, инициируемых объектами a_i, опишем как $E = \{e_{111}, e_{112}, ..., e_{ij_ik_{ij_i}}, e_{IJ_iK_{ij_i}}\}$, где $e_{ij_ik_{ij_i}}$ – событие, для которого вектор V_{ij_i} имеет вид: $V_{ij_i} = [0,0,..., d_{ij_ik_{ij_i}} = 1,...,0]$ или в отношении функции $c_{ij_ik_{ij_i}}$ можно сделать вывод: $c_{ij_ik_{ij_i}}(b_{ij_i}) = 1$.

Общее количество событий $e_{ij_ik_{ij_i}} \in E$ равно:

$$N^E = \sum_{i=1}^{I} \sum_{j=1}^{J_i} K_{ij_i} . \qquad (9)$$

Выделим отдельно подмножество $E(s_1) \subset E$, элементами которого являются события, представляющие собой факт принятия того или иного значения параметрами $B(s_1)$.

Следует заметить, что генерация событий происходит не на стороне SNMP-«агента», а в вычислительной среде модуля системы диагностики, который реализует функции «менеджера». Как показано на рис. 2, модуль периодически и поэтапно выполняет следующие действия:

– отправляет запрос SNMP-«агентам» на получение текущих значений параметров $b_{ij_i} \in B_i$;

– вычисляет значение функций $c_{ij_i k_{ij_i}}(b_{ij_i}) \in C_{ij_i}$ для полученных на первом этапе значений параметров $b_{ij_i} \in B_i$;

– формирует вектор V_{ij_i} и на основе его данных генерирует событие $e_{ij_i k_{ij_i}} \in E$.

Обозначим моменты времени начала выполнения модулем перечисленных выше действий в виде множества $T = \{t_1, t_2, ..., t_p, t_P\}, p = \overline{1,P}$, где P – количество раз выполнения модулем перечисленных выше этапов. В каждый момент времени t_p сгенерированные на третьем этапе события формируют множество $E_p \subset E$.

Определение 7. Множество событий $E_p \subset E$ будем называть описанием состояния IP-сети в момент времени t_p.

Определение 8. Множество событий $E_p(s_l) = E_p \cap E(s_l)$ будем называть описанием состояния службы s_l в момент времени t_p.

Очевидно, что все события $e_{ij_i k_{ij_i}} \in E_p$ имеют одинаковое время возникновения, а общее количество состояний IP-сети, которые можно описать, используя изложенную выше событийно-ориентированную модель, равно:

$$N^{\Omega} = \prod_{i=1}^{I} \prod_{j=1}^{J_i} K_{ij_i}. \qquad (10)$$

Определение 9. Базой данных диагностической информации (БДДИ) называется набор Ψ записей вида $\psi_p = (E_p, p)$, где E_p – совокупность событий, которые описывают значения диагностических параметров в момент времени p.

Определение 10. Ситуационной моделью IP-сети для целей формирования БДДИ называется модель, представляемая объектами:

$$M_{co} = \left\langle S, A\{B_i\{C_{ij_i}, D_{ij_i}, V_{ij_i}, \{e_{ij_i k_{ij_i}}\}_{k_{ij_i}}^{K_{ij_i}}\}_{j_i=1}^{J_i}\}_{i=1}^{I}, T, \Psi \right\rangle. \qquad (11)$$

3. Выводы

Обобщим приведенные в статье выкладки в виде следующих ключевых особенностей модели (11):

– IP-сеть представляется в виде конечного множества служб S и ИДИ A. Каждый ИДИ $a_i \in A$ в свою очередь является совокупностью доступных для контроля параметров $b_{ij_i} \in B_i$.

– Событие $e_{ij_i k_{ij_i}} \in E$ есть факт принятия параметром b_{ij_i} ИДИ a_i значения из множества $D_{ij_i k_{ij_i}} \subset D(b_{ij_i})$ или, используя аппарат характеристических функций, факт принятия функцией $c_{ij_i k_{ij_i}}(b_{ij_i})$ единичного значения.

– События генерируются в строго определенные моменты времени $t_p \in T$ и формируют множество $E_p \subset E$.

– Состояние IP-сети в момент времени $t_p \in T$ описывается множеством $E_p \subset E$.

– Состояние службы s_l в момент времени $t_p \in T$ описывается множеством $E_p(s_l) \subset E(s_l)$.

– На заданном временном интервале $(t_p - t_1)$ транзакции $\psi_p = (E_p, p)$ формируют БДДИ IP-сети.

Научная новизна исследования состоит в разработке новой диагностической модели IP-сети (11), которая позволяет сформировать БДДИ для дальнейшего применения методов Data Mining в целях выявления закономерностей в процессе функционирования аппаратных и программных компонентов IP-сети.

Практическая ценность предложенной диагностической модели вытекает из возможности ее использования в процессе разработки методики определения состояния IP-сети и ее служб, а также новых методов диагностирования, в основе которых лежит применение подходов искусственного интеллекта и Data Mining.

Список литературы: 1. *Ермаков А.А.* Основы надежности информационных систем: учебное пособие. Иркутск: ИрГУПС, 2006. 151с. **2.** *Черкесов Г.Н.* Надежность аппаратно-программных комплексов. Учебное пособие. СПб.: Питер, 2005. 479с. **3.** *Coates M., Hero A., Nowak R., Yu B.* Internet Tomography / / IEEE Signal Processing Magazine, May 2002. **4.** *Caceres R., Duffield N.G., Horowitz J., Towsley D.* Multicast-based inference of network-internal loss characteristics. IEEE Transactions on Information Theory 45 (1999) 2462–2480 toit99.pdf. **5.** *Adams A.* The Use of End-to-end Multicast Measurements for Characterizing Internal Network Behavior / A. Adams, T. Bu, R. Caceres, N. Duffield, J. Horowitz, F. Lo Presti, S. B. Moon, V. Paxson, D. Towsley – CiteSeer. Scientific Literature Digital Library and Search Engine. – Режим доступа: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.2318. **6.** *Katzela I.* Schemes for fault identification in communication networks [Текст] / I.Katzela, M. Schwartz // IEEE/ACM Transactions on Networking. – 1995. Vol.3. P.753—764. katzela95schemes.pdf. **7.** *Кучер А.В.* Динамический анализ и диагностика состояния IP-сети [Электронный ресурс] : дис. … канд. техн. наук / А.В.Кучер. М.: РГБ, 2007 (Из фондов Российской Государственной библиотеки). Режим доступа: http://diss.rsl.ru/diss/05/0616/050616043.pdf. **8.** *Поморова О.В.* Теоретичні основи, методи та засоби інтелектуального діагностування комп'ютерних систем [Текст] : автореф. дис. … д-ра техн. наук : 05.13.13 / О.В.Поморова; [Національний університет «Львівська політехніка»]. Львів, 2007. 29с. **9.** *Klemettinen M.* Rule discovery in telecommunication alarm data [Текст] / M. Klemettinen, H. Mannila, H. Toivonen // Journal of Network and Systems Management. 1999. Vol.7(4). P. 395-423. 10.1.1.7.7471.pdf.

Поступила в редколлегию 22.05.2009

Соколов Сергей Алексеевич, канд. техн. наук, профессор, зав. кафедрой Харьковского университета Воздушных Сил. Научные интересы: обработка информации в телекоммуникационных системах. Адрес: Украина, 61000, Харьков, тел. 8-577-342-22-84.

Стокипный Александр Леонидович, офицер отдела связи и автоматизации Восточного регионального управления Государственной пограничной службы Украины, соискатель ХУПС. Научные интересы: применение методов ИИ в современных телекоммуникационных системах. Адрес: Украина, 61000, Харьков, ул. Героев Труда, 46, кв.183, тел. 8-067-573-19-16.

Голдаев Алексей Витальевич, студент 5 курса факультета Компьютерной инженерии и управления ХНУРЭ. Адрес: Украина, 61000, Харьков, ул. Гвардейцев Широнинцев, 81, кв. 48, тел. 8-099-066-31-44.