Практическая значимость результатов исследования заключается в том, что использование результатов комплекса задач обеспечивает реальное повышение эффективности эксплуатации средств ЭХЗ магистральных газопроводов, позволяет сократить количество отказов трубопроводов и сократить затраты электроэнергии на УКЗ.

Литература

1. ДСТУ 4219:2003 Трубопроводи сталеві магістральні. Загальні вимоги до захисту від корозії
2. Бэкман В., Швенк В. Катодная защита от коррозии: Справ, изд. Пер. с нем. - М.: Металлургия, 1984. - 496 с.
3. СТП 320.30019801.072-2003 Магістральні газопроводи. Розрахунок електрохімічного захисту
4. Тевяшев А.Д., Ткаченко В.Ф., Попов А.В., Стрижак Л.В. Стохастический поход к постановке и решению задачи оперативного планирования режима работы системы ЭХЗ трубопровода //Восточно-Европейский журнал передовых технологий. 2005. №15 с. 94-98.

Дано опис метода діагностування корпоративної IP-мережі, в якому реалізована концепція інтелектуального аналізу даних. В якості джерела даних пропонується використовувати базу даних діагностичної інформації, яка містить опис роботи компонентів IP- мережі у вигляді набору подій

Ключові слова: IP-мережа, методи діагностики, асоціативні правила, інтелектуальний аналіз даних

Описан метод диагностирования корпоративной IP-сети, в котором реализована концепция интеллектуального анализа данных. В качестве источника данных предлагается использовать базу данных диагностической информации, которая содержит описания работы компонентов IP-сети в виде событий

Ключевые слова: IP-сеть, методы диагностики, ассоциативные правила, интеллектуальный анализ данных

The method of diagnostics of corporate IP networks, conception of intellectual analysis of data is realized in which, is described. As a source of data it is suggested to use the database of diagnostic information which contains descriptions of work of components of IP networks as events

Keywords: IP- networks, methods of diagnostics, associative rules, intellectual analysis of data

УДК 681.003.66

МЕТОД ПОИСКА ЗАКОНОМЕРНОСТЕЙ В БАЗЕ ДАННЫХ ДИАГНОСТИЧЕСКОЙ ИНФОРМАЦИИ КОРПОРАТИВНОЙ IP-СЕТИ

А.Л. Стокипный
Офицер отдела
Восточное региональное управление Государственной пограничной службы Украины
г. Харьков
Контактний тел.: 8 (057) 700-92-06
E-mail — a.stokipny@gmail.com

1. Введение

На сегодняшний день в ходе создания подавляющего большинства современных корпоративных информационно-телекоммуникационных систем применяют IP-технологии и протоколы. Корпоративная IP-сеть является основой функционирования ключевых компонентов современных информационных систем – пользовательских служб, распределенный характер которых выдвигает высокие требования к доступности используемой сети передачи данных. Повысить доступность корпоративной IP-сети можно путем сокращения времени простоя, составляющими которого с одной стороны является время, необходимое для проведения плановых работ по техническому обслуживанию, а с другой – время, затрачиваемое об-

служивающим персоналом для восстановления работоспособного состояния корпоративной IP-сети после возникновения различного рода неисправностей. Применение резервирования отдельных компонентов IP-сети дает возможность сократить время простоя, однако является достаточно дорогостоящей процедурой. Наличие комплексной системы диагностики, в состав которой входят эффективные методы диагностирования, позволяет сократить время простоя корпоративной IP-сети, связанное с поиском причин возникновения неисправностей, их локализацией и проведением адекватных восстановительных мероприятий. Анализ литературы показал, что в ходе создания систем диагностики все чаще применяют подходы, основанные на знаниях [1,2,3]. Причем знания могут быть, как экспертного типа, так и полученные в ходе анализа баз данных диагностической информации (БДДИ), которые содержат описания работы компонентов IP-сети в виде значений диагностических параметров за длительный период времени. Целесообразность разработки систем диагностики, основанных на знаниях, подтверждаются также в ходе исследований, основным направлением которых является поиск методов автоматизации процесса создания и модификации базы знаний, сформированной на основе анализа диагностической информации [4,5,6].

Целью статьи будем считать разработку метода поиска закономерностей в БДДИ корпоративной IP-сети, который может быть в дальнейшем использован для создания и поддержки в актуальном состоянии базы знаний комплексной системы диагностики IP-сети.

2. Диагностическая модель корпоративной IP-сети

Предлагается формировать БДДИ корпоративной IP-сети в соответствии со следующей моделью:

$$M=\left\langle A\left\{B_i\left\{D_{ij_i},\left\{e_{ij_ik_{ij_i}}\right\}_{k_{ij_i}}^{K_{ij_i}}\right\}_{j_i=1}^{J_i}\right\}_{i=1}^{I},T,\Psi\right\rangle \qquad (1)$$

где A – множество источников диагностической информации (ИДИ), $|A|=I$; B_i – множества диагностических параметров ИДИ a_i, $|B_i|=J_i$; D_{ij_i} – множество интервалов, на которые разбита область значений параметра b_{ij_i}, $|D_{ij_i}|=K_{ij_i}$; $e_{ij_ik_{ij_i}}$ – событие, состоящее в принятии параметром b_{ij_i} значения из множества $D_{ij_ik_{ij_i}}$; T – моменты времени генерации событий $e_{ij_ik_{ij_i}}$ (опрос ИДИ и получение значений диагностических параметров); Ψ – упорядоченный набор записей вида $\psi_p=(E_p,p)$; E_p – совокупность событий, сгенерированных в момент времени $p\in T$.

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 (1), поскольку на сегодняшний день подавляющее большинство аппаратных и программных компонентов содержит реализацию указанного протокола.

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

Определение 1. Частным диагностическим параметром (ЧДП) называется параметр, который отражает индивидуальную характеристику отдельно взятого аппаратного или программного компонента.

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

Определение 2. Общим диагностическим параметром (ОДП) называется параметр, который содержит агрегированную оценку функционирования системы взаимодействующих компонентов.

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

IP-сеть обеспечивает работу множества служб $S=\{s_1,s_2\ldots s_l\ldots s_L\},l=\overline{1,L}$, где L – общее количество служб.

3. Методика определения состояния службы IP-сети

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 3. Собственным параметром службы s_l называется диагностический параметр b_{ij_i}, на множестве значений $D(b_{ij_i})$ которого задано разбиение $D_{ij_ik^0{}_{ij_i}}(b_{ij_i})\cup D_{ij_ik^1{}_{ij_i}}(b_{ij_i})$ такое, что значения из $D_{ij_ik^0{}_{ij_i}}(b_{ij_i})$ соответствуют работоспособному состоянию службы s_l, а значения из $D_{ij_ik^1{}_{ij_i}}(b_{ij_i})$ – неработоспособному.

Определение 4. Диагностический параметр не являющийся собственным параметром службы s_l, называется неопределенным параметром службы s_l.

Относительно элементов множества $B(s_l)$ следует заметить, что использование их для идентификации состояния службы s_l, предполагает выполнение важного условия: для каждого параметра $b_{ij_i}\in B(s_l)$ известны множества значений, которые он принимает в случае работоспособного и неработоспособного состояний службы s_l, следовательно, каждый из параметров $b_{ij_i}\in B(s_l)$ является собственным параметром службы s_l.

Определение 5. Стандартным (номинальным) значением параметра $b_{ij_i}\in B(s_l)$ относительно службы s_l называется значение, которое принимает параметр $b_{ij_i}\in B(s_l)$ в случае, если служба s_l находится в работоспособном состоянии.

В модели (1) стандартные значения собственных диагностических параметров в обозначены как $D_{ij_ik^0{}_{ij_i}}(b_{ij_i})$, а множество значений, соответствующих неработоспособному состоянию службы s_l – как $D_{ij_ik^1{}_{ij_i}}(b_{ij_i})$.

Из определения неисправности в [7] следует, что под неисправностью следует понимать событие, при котором происходит недопустимое отклонение значений параметра системы от стандартного (номинального) значения. То есть событие $e_{ij_i k^1_{ij_i}} \in E$, состоящее в принятии параметром $b_{ij_i} \in B(s_l)$ значения из множества $D_{ij_i k^1_{ij_i}}(b_{ij_i})$, является неисправностью и свидетельствует о факте перехода службы s_l в неработоспособное состояние. Следовательно, для некоторой службы s_l существует одно работоспособное состояние $\Omega^p(s_l)$, при котором значение любого из параметров $b_{ij_i} \in B(s_l)$ принадлежит множествам $D_{ij_i k^0_{ij_i}}(b_{ij_i})$, и конечное множество неработоспособных состояний $\Omega^{np}(s_l) = \left\{\Omega^{np}_1(s_l), \Omega^{np}_2(s_l) ... \Omega^{np}_{\tau_l}(s_l), \Omega^{np}_{N^{\Omega_{np}}_{\tau_l}}(s_l)\right\}$, $\tau_l = \overline{1..N^{\Omega_{np}}_{\tau_l}}$, каждое из которых характеризуются наличием хотя бы одного параметра $b_{ij_i} \in B(s_l)$ со значением, принадлежащим множеству $D_{ij_i k^1_{ij_i}}(b_{ij_i})$.

Общее количество неработоспособных состояний службы s_l равно $N^{\Omega_{np}}_{\tau_l} = 2^{B(s_l)} - 1$.

Расширим модель (1) и дополнительно введем обозначения:

$E^0(s_l)$ – совокупность вариантов формирования множества $E_p(s_l)$, при котором $E_p(s_l)$ не содержит ни одной неисправности.

$E^1(s_l)$ – совокупность вариантов формирования множества $E_p(s_l)$, при которых $E_p(s_l)$ содержит хотя бы одну неисправность.

$E^*(s_l) = E^0(s_l) \cup E^1(s_l)$ – совокупность всех возможных вариантов формирования множества $E_p(s_l)$.

Методику определения технического состояния некоторой службы s_l в терминах модели (1) определим следующим образом:

1. Задается множество собственных параметров $B(s_l)$, которые будут использоваться в процессе идентификации состояния s_l.

2. Для каждого $b_{ij_i} \in B(s_l)$ на множестве значений $D(b_{ij_i})$ определяется два подмножества таким образом, чтобы значения подмножества $D_{ij_i k^0_{ij_i}}(b_{ij_i}) \subset D(b_{ij_i})$ соответствовали работоспособному состоянию службы s_l, а значения $D_{ij_i k^1_{ij_i}}(b_{ij_i}) \subset D(b_{ij_i})$ – неработоспособному.

3. Формируется множество $E^*(s_l)$ и выделяются его подмножества $E^0(s_l)$ и $E^1(s_l)$. Исходя из того, что для идентификации состояния службы s_l используются только собственные параметры (этап 1) и каждый параметр идентифицирует два возможных состояния службы (этап 2), тогда множество $E^0(s_l)$ содержит единственный элемент, который представляет собой совокупность событий $e_{ij_i k^0_{ij_i}}$.

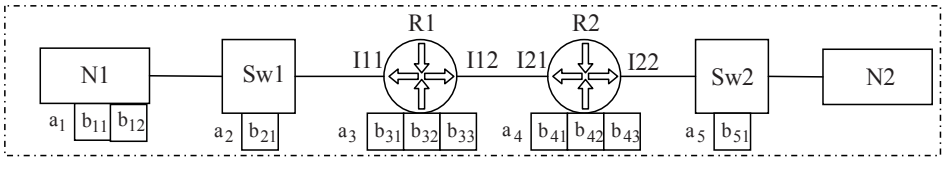
4. Работоспособному состоянию $\Omega^p(s_l)$ ставится в соответствие элемент множества $E^0(s_l)$.

5. Задается $N^{\Omega_{np}}_{\tau_l} = |E^1(s_l)|$ доступных для идентификации неисправных состояний. Определяется биективная функция $f: E^1(s_l) \to \Omega^{np}(s_l)$ и обратная ей $f^{-1}: \Omega^{np}(s_l) \to E^1(s_l)$. Результатом вычисления $f^{-1}(\Omega^{np}_{\tau_l}(s_l))$ есть множество событий, на основе которых происходит идентификация состояния $\Omega^{np}_{\tau_l}(s_l) = f(E^1(s_l))$.

6. Техническое состояние службы s_l в момент времени t_p равно:

– $\Omega^p(s_l)$, если работоспособна или формально, если $E_p(s_l) \in E^0(s_l)$;

– $f(E_p(s_l))$, если служба неработоспособна или формально, если $E_p(s_l) \in E^1(s_l)$.



b_{11} – время односторонней задержки доставки пакета фиксированной длины от узла N1 к узлу N2.

b_{12} – скорость передачи данных между узлами N1 и N2.

b_{21} – процент использования пропускной способности порта коммутатора Sw1, к которому подключен маршрутизатор R1.

b_{31} – скорость появления пакетов с ошибками на интерфейсе I11 маршрутизатора R1.

b_{32} – процент загрузки процессора маршрутизатора R1.

b_{33} – скорость появления пакетов с ошибками на интерфейсе I12 маршрутизатора R1.

b_{41} – скорость появления пакетов с ошибками на интерфейсе I21 маршрутизатора R2.

b_{42} – процент загрузки процессора маршрутизатора R2.

b_{43} – скорость появления пакетов с ошибками на интерфейсе I22 маршрутизатора R2.

b_{51} – процент использования пропускной способности порта коммутатора Sw2, к которому подключен маршрутизатор R2.

Рис. 1. Схема построения типового участка IP-сети

Рассмотрим процесс диагностирования службы IP-телефонии на участке IP-сети, схема которого представлена на рис.1. Для доступа к голосовому шлюзу N2 используется канал передачи данных, образованный коммутаторами «Sw1» и «Sw2», маршрутизаторами «R1», «R2» и линиями связи. Указанное оборудование содержит SNMP-агентов, обеспечивающих доступ к текущим значениям SNMP-объектов управления посредством соответствующего протокола. Рабочая станция N1 предназначена для сбора и анализа диагностических данных, поступающих через протокол SNMP. Кроме того, программное обеспечение N1 позволяет измерить характеристики канала передачи данных на участке «N1-Sw1-R1-R2-Sw2-N2» и предоставить текущие значения в виде SNMP-объектов управления.

Процесс диагностирования службы IP-телефонии на участке IP-сети (рис. 1), опишем в виде диагностической модели (1). Основные объекты модели и их описание представлены в таблице 1.

Таблица 1

Диагностическая модель участка IP-сети (рис.1)

Объект модели	Описание
$S=\{s_1\}, L=1$	s_1 – служба IP-телефонии
$A:=\{a_1,a_2,a_3,a_4,a_5\}, I=5$	a_1 – рабочая станция N1, a_2 – коммутатор Sw1 a_3 – маршрутизатор R1, a_4 – маршрутизатор R2 a_5 – коммутатор Sw2
$B_1=\{b_{11},b_{12}\}, J_1=2$	SNMP-объекты для контроля общих характеристик канала передачи данных «N1-Sw1-R1-R2-Sw2-N2»
$B_2=\{b_{21}\}, J_2=1$	Диагностические параметры, соответствующие SNMP-объектам коммутатора Sw1
$B_3=\{b_{31},b_{32},b_{33}\}, J_3=3$	Диагностические параметры, соответствующие SNMP-объектам маршрутизатора R1
$B_4=\{b_{41},b_{42},b_{43}\}, J_4=3$	Диагностические параметры, соответствующие SNMP-объектам маршрутизатора R2
$B_5=\{b_{51}\}, J_5=1$	Диагностические параметры, соответствующие SNMP-объектам коммутатора Sw2
$N^B=10$	Общее количество диагностических параметров
$D(b_{11})=\{1...400\}$, $D_{111}(b_{11})=\{1...200\}$ $D_{112}(b_{11})=\{201...400\}$, $K_{11}=2$	Рекомендуемые значения времени односторонней задержки доставки пакета для различных типов каналов передачи данных содержатся в документе ITU-T G.114 “One way transmission time”. Значения в миллисекундах.
$D(b_{12})=\{1...2000\}$, $D_{121}(b_{12})=\{41...2000\}$ $D_{112}(b_{11})=\{1...40\}$, $K_{12}=2$	Рекомендуемые значения скорости передачи данных для различных типов кодеков определены на основе рекомендаций, изложенных в документе ITU-T G.113 “Transmission impairments”. Значения в Кбит/секунду.
$D(b_{21})=D(b_{32})=D(b_{42})=D(b_{51})=\{1...100\}$ $D_{221}(b_{21})=D_{321}(b_{32})=D_{421}(b_{42})=D_{511}(b_{51})=\{1...20\}$ $D_{212}(b_{21})=D_{322}(b_{32})=D_{422}(b_{42})=D_{512}(b_{51})=\{21...40\}$ $D_{213}(b_{21})=D_{323}(b_{32})=D_{423}(b_{42})=D_{513}(b_{51})=\{41...60\}$ $D_{214}(b_{21})=D_{324}(b_{32})=D_{424}(b_{42})=D_{514}(b_{51})=\{61...80\}$ $D_{215}(b_{21})=D_{325}(b_{32})=D_{425}(b_{42})=D_{515}(b_{51})=\{81...100\}$ $D(b_{31})=D(b_{33})=D(b_{41})=D(b_{43})=\{1...500\}$ $D_{311}(b_{31})=D_{331}(b_{33})=D_{411}(b_{41})=D_{431}(b_{43})=\{1...100\}$ $D_{312}(b_{31})=D_{332}(b_{33})=D_{412}(b_{41})=D_{432}(b_{43})=\{101...200\}$ $D_{313}(b_{31})=D_{333}(b_{33})=D_{413}(b_{41})=D_{433}(b_{43})=\{201...300\}$ $D_{314}(b_{31})=D_{334}(b_{33})=D_{414}(b_{41})=D_{434}(b_{43})=\{301...400\}$ $D_{315}(b_{31})=D_{335}(b_{33})=D_{415}(b_{41})=D_{435}(b_{43})=\{401...500\}$ $K_{21}=K_{31}=K_{32}=K_{33}=K_{41}=K_{42}=K_{43}=K_{51}=5$	Относительно SNMP-объектов, представленных диагностическими параметрами b_{21}, b_{31}, b_{32}, b_{33}, b_{41}, b_{42}, b_{43}, b_{51}, отсутствует информация о том, каким образом их значения распределены между работоспособным и неработоспособным состояниями службы s_1. С целью выявления указанного распределения, множество значений параметров равномерно разбивается на подмножества для проведения дальнейшего исследования.
$B(s_1)=\{b_{11},b_{12}\}$	Согласно вышеизложенным требованиям, из всех доступных для контроля диагностических параметров в состав множества $B(s_1)$ могут входить только b_{11} и b_{12}, поскольку для остальных восьми неизвестны значения, соответствующие работоспособному и неработоспособному состоянию службы s_1.
$E=\{e_{111}, e_{112}, e_{121}, e_{122}, e_{211}, e_{212}, e_{213}, e_{214}, e_{215}, e_{311}, e_{312}, e_{313}, e_{314}, e_{315}, e_{321}, e_{322}, e_{323}, e_{324}, e_{325}, e_{331}, e_{332}, e_{333}, e_{334}, e_{335}, e_{411}, e_{412}, e_{413}, e_{414}, e_{415}, e_{421}, e_{422}, e_{423}, e_{424}, e_{425}, e_{431}, e_{432}, e_{433}, e_{434}, e_{435}, e_{511}, e_{512}, e_{513}, e_{514}, e_{515}\}$ $N^E=44$	Множество событий, которые могут использоваться для идентификации состояния корпоративной IP-сети.
$E(s_1)=\{e_{111}, e_{112}, e_{121}, e_{122}\}$	События, которые могут использоваться для описания состояния службы s_1.
$N^\Omega=1562500$	Количество состояний IP-сети (рис.1), которые можно описать с помощью событий из множества E.
$E_p=\{e_{11k_{11}^p}, e_{12k_{12}^p}, e_{31k_{31}^p}, e_{32k_{32}^p}, e_{42k_{42}^p}, e_{43k_{43}^p}, e_{51k_{51}^p}\}$ $\psi_p=(E_p,p)$	ψ_p - запись БДДИ, которая описывает состояние IP-сети (рис.1) в момент времени $t_p \in T$.

Продолжение таблицы 1

$E_p(s_l)=\left\{e_{11k_{11}^p},e_{12k_{12}^p}\right\}$	Описание состояния службы s_1 в момент времени $t_p \in T$.
$E^0(s_l)=\{\{e_{111},e_{121}\}\}$	Вариант формирования $E_p(s_l)$, соответствующий работоспособному $\Omega^p(s_l)$ состоянию службы s_1.
$E^1(s_l)=\{\{e_{111},e_{122}\},\{e_{112},e_{121}\},\{e_{112},e_{122}\}\}$	Варианты формирования $E_p(s_l)$, соответствующие неработоспособным состояниям службы s_1.
$f(\{e_{111},e_{122}\})\to\Omega_1^{np}(s_l)$, $f(\{e_{112},e_{121}\})\to\Omega_2^{np}(s_l)$ $f(\{e_{112},e_{122}\})\to\Omega_2^{np}(s_l)$	$f:E^1(s_l)\to\Omega^{np}(s_l)$ – функция для идентификации неисправных состояний службы s_1.

Допустим, что в момент времени $t_1<t_{p'}<t_P$ часть компьютеров, подключенных к коммутатору Sw2, была заражена вирусом, который начал генерировать избыточный трафик, проходящий через канал «Sw1-R1-R2-Sw2».

Функционирование вируса стало причиной изменений следующих характеристик работы сетевого оборудования:

– увеличение количества IP-пакетов, проходящих через порт коммутатора Sw1, к которому подключен маршрутизатор R1;

– увеличение процессорного времени, необходимого для обработки маршрутизаторами R1 и R2 поступающих на них пакетов;

– увеличение количества IP-пакетов, проходящих через порт коммутатора Sw2, к которому подключен маршрутизатор R2.

Также изменились и перестали соответствовать требованиям службы s_l значения общих характеристик канала передачи данных:

– увеличилось время односторонней задержки доставки пакета фиксированной длины от узла N1 к узлу N2;

– уменьшилась скорость передачи данных между узлами N1 и N2.

Симптомами проявления функционирования вируса являются описанные выше изменения значений параметров b_{11}, b_{12}, b_{21}, b_{32}, b_{42}, b_{51}, однако в процессе определения состояния службы s_l используются только параметры b_{11}, b_{12}, которые являются ОДП и выражают требования службы не к характеристикам отдельных компонентов, а к общим характеристикам канала передачи данных как системы взаимодействующих компонентов. Вследствие этой особенности применение ОДП при поиске причин неисправности неэффективно, поскольку по значениям b_{11}, b_{12} мы не можем заключить, какие из характеристик компонентов отклонились от значений работоспособного состояния службы, и, следовательно, не можем провести локализацию неисправности с точностью до характеристики отдельного компонента.

В рассматриваемой ситуации дополнительные данные для выявления причин отклонения параметров b_{11}, b_{12} от стандартных значений доступны посредством ЧДП b_{21}, b_{32}, b_{42}, b_{51} компонентов Sw1, R1, R2, Sw2. Однако применение указанных ЧДП невозможно, так как они являются неопределенными параметрами службы s_1 и для каждого из них неизвестны значения, соответствующие работоспособному и неработоспособному состоянию службы s_1.

Для использования неопределенных параметров в процессе поиска причин неисправного состояния службы необходимо разработать метод, который бы посредством анализа БДДИ, содержащей в событийно-ориентированном виде результаты наблюдения за работой корпоративной IP-сети, позволял находить закономерности между фактами принятия значений параметрами из множества $B(s_l)$ и параметрами из множества $\left(B=\bigcup_i^I B_i\right)/B(s_1)$.

4. Метод поиска закономерностей в БДДИ корпоративной IP-сети.

Поскольку событие $e_{ij_ik_{ij_i}}$ представляет собой факт принятия параметром b_{ij_i} значения из множества $D_{ij_ik_{ij_i}}$, тогда поиск закономерностей между значениями собственных параметров из множества $B(s_l)$ и значениями неопределенных параметров из множества $\left(B=\bigcup_i^I B_i\right)/B(s_1)$ равносильно поиску закономерностей между возникновением событий из множества $E_p(s_l)$ и событий из множества $E_p/E_p(s_l)$.

Будем считать, что в момент времени $t_{p'-1}$ (до перехода в неработоспособное состояние), служба s_l находилась в работоспособном состоянии $\Omega^p(s_l)$, описанием которого являлись наборы событий $E_{p'-1}(s_l) = \{e_{111},e_{121}\}$ и $E_{p'-1}/E_{p'-1}(s_l)=\left\{e_{21k_{21}^{p'-1}}, e_{31k_{31}^{p'-1}},e_{32k_{32}^{p'-1}},e_{33k_{33}^{p'-1}}, e_{41k_{41}^{p'-1}},e_{42k_{42}^{p'-1}},e_{43k_{43}^{p'-1}},e_{51k_{51}^{p'-1}}\right\}$.

Из этого следует, что неопределенные параметры $b_{21}, b_{31}, b_{32}, b_{33}, b_{41}, b_{42}, b_{43}, b_{51}$ в момент времени $t_{p'-1}$ приняли значения из множеств $D_{21k_{21}^{p'-1}}$, $D_{31k_{31}^{p'-1}}$, $D_{32k_{32}^{p'-1}}$, $D_{33k_{33}^{p'-1}}$, $D_{41k_{41}^{p'-1}}$, $D_{42k_{42}^{p'-1}}$, $D_{43k_{43}^{p'-1}}$, $D_{51k_{51}^{p'-1}}$.

Вследствие функционирования вируса в момент времени $t_{p'}$ служба перешла в неработоспособное состояние $\Omega_2^{np}(s_l)=f(\{e_{112},e_{122}\})$, которому соответствуют наборы событий $E_{p'}(s_l)=\{e_{112},e_{122}\}$ и $E_{p'}/E_{p'}(s_l)=\left\{e_{21k_{21}^{p'}},e_{31k_{31}^{p'-1}},e_{32k_{32}^{p'}},e_{33k_{33}^{p'-1}}, e_{41k_{41}^{p'-1}},e_{42k_{42}^{p'}},e_{43k_{43}^{p'-1}},e_{51k_{51}^{p'}}\right\}$.

Следовательно, значения неопределенных параметров b_{21}, b_{32}, b_{42}, b_{51} изменились таким образом, что стали принадлежать множествам $D_{21k_{21}^{p'}}$, $D_{32k_{32}^{p'}}$, $D_{42k_{42}^{p'}}$, $D_{51k_{51}^{p'}}$.

Анализируя состав множества $E_{p'}/E_{p'}(s_1)$ можно выделить две группы неопределенных параметров:

– Независимые неопределенные параметры (ННП), которые изменяют свои значения независимо от состояния службы s_1. В рассматриваемом примере к ННП относятся параметры (b_{31}, b_{33}, b_{41}, b_{43}), значения которых в случае пребывания службы s_1 как в работоспособном, так и в неработоспособном состоянии принадлежат одному и тому же множеству значений.

– Зависимые неопределенные параметры (ЗНП), изменение значений которых в совокупности или по отдельности является индикатором смены состояния службы. В рассматриваемом примере к группе ЗНП относятся параметры (b_{21}, b_{32}, b_{42}, b_{51}), значения которых при переходе службы в неработоспособное состояние изменяются и перестают соответствовать значениям, характерным для ее работоспособного состояния.

Исходя из определения ННП, особенности функционирования компонентов, отраженные в значениях ННП, не могут быть причиной перехода службы в неработоспособное состояние. Следовательно, учет их значений при поиске причин неработоспособного состояния службы является избыточной процедурой, что позволяет нам исключить группу ННП из дальнейшего рассмотрения и принимать во внимание только значения ЗНП.

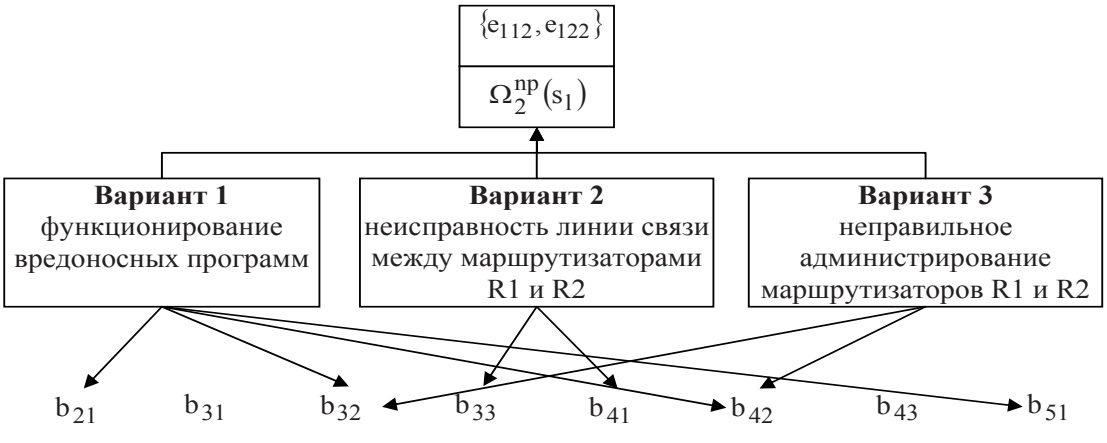


Рис. 2. Схема соответствия ЗДП различным вариантам причин возникновения неработоспособного состояния $\Omega_2^{пр}(s_1)$ службы s_1.

Некоторому неработоспособному состоянию службы может соответствовать несколько наборов ЗНП. Следует заметить, что принадлежность параметра к ЗНП или ННП зависит только от причин, которые вызвали неисправное состояние службы, и позволяет классифицировать параметры только относительно

Таблица 2

Описание причин возникновения неработоспособного состояния $\Omega_2^{пр}(s_1)$ службы s_1.

№ варианта	ЗНП	ННП	Описание
1	b_{21}, b_{32}, b_{42}, b_{51}	b_{31}, b_{33}, b_{41}, b_{43}	Данный вариант рассматривается в качестве примера, его детальное описание представлено ранее в статье.
2	b_{33}, b_{41}	b_{21}, b_{31}, b_{32}, b_{42}, b_{43}, b_{51}	Канал передачи данных между маршрутизаторами R1 и R2 построен с использованием xDSL модемов, которые работают в однопарном режиме на двух медных линиях связи. Повреждение одной из линий стало причиной увеличения количества фреймов, переданных/принятых с ошибкой на канальном уровне. Необходимость повторной передачи повлияла как на характеристики канала «R1-R2», так и канала «Sw1-R1-R2-Sw2».
3	b_{32}, b_{42}	b_{31}, b_{32}, b_{33}, b_{41}, b_{43}, b_{51}	Маршрутизаторы R1 и R2 представляют собой устройства Cisco 3800 series. В результате изменения конфигурации был включен режим протоколирования IP-трафика. Выполнение данной задачи чрезвычайно требовательно к ресурсам процессора, вследствие чего процент загрузки процессора поднялся до 95% (b_{21}). Даже при нормальной интенсивности трафика это привело к появлению задержек и снижению скорости передачи на всем участке канала «Sw1-R1-R2-Sw2».

заданного неисправного состояния. Данный факт проиллюстрирован на рис. 2 в виде нескольких вариантов причин возникновения неработоспособного состояния $\Omega_2^{np}(s_l)$.

Комментарии к каждому из вариантов приведены в таблице 2.

Совокупность наборов событий, которые отражают состояния ЗДП для некоторого неработоспособного состояния, представим в виде множества

$$X_{l\tau_l} = \left\{x_{l\tau_l 1}, x_{l\tau_l 2}, ..., x_{l\tau_l g_{l\tau_l}}, ..., x_{l\tau_l G_{l\tau_l}}\right\}, g_{l\tau_l} = \overline{1, G_{l\tau_l}},$$

где $G_{l\tau_l}$ – общее количество наборов, $x_{l\tau_l g_{l\tau_l}}$ – набор событий с порядковым номером $g_{l\tau_l}$, который соответствует неработоспособному состоянию $\Omega_{\tau_l}^{np}(s_l)$ службы s_l.

В случае возникновения одной из трех описанных в таблице 2 ситуаций, неработоспособному состоянию $\Omega_2^{np}(s_l) = f\left(\left\{e_{112}, e_{122}\right\}\right)$ соответствуют следующие наборы событий:

1. $x_{111} = \left\{e_{21k_{11}, k_{11} \neq k_{11}^{p'-1}}, e_{32k_{32}, k_{32} \neq k_{32}^{p'-1}}, e_{42k_{42}, k_{42} \neq k_{42}^{p'-1}}, e_{51k_{51}, k_{51} \neq k_{51}^{p'-1}}\right\}$;
2. $x_{112} = \left\{e_{33k_{33}, k_{33} \neq k_{33}^{p'-1}}, e_{41k_{41}, k_{41} \neq k_{41}^{p'-1}}\right\}$;
3. $x_{113} = \left\{e_{21k_{21}, k_{21} \neq k_{21}^{p'-1}}, e_{42k_{42}, k_{42} \neq k_{42}^{p'-1}}\right\}$.

Искомую закономерность между возникновением событий из $E_p(s_l)$ и событий из $E_p / E_p(s_l)$ обозначим в виде ассоциативного правила $\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}$.

Определение 6.

Набор событий $x_{l\tau_l g_{l\tau_l}}$ имеет поддержку $\sup(x_{l\tau_l g_{l\tau_l}})$, если $\left(\sup(x_{l\tau_l g_{l\tau_l}}) * P\right)$ записей БДДИ корпоративной IP-сети содержат $x_{l\tau_l g_{l\tau_l}}$.

Определение 7.

Ассоциативное правило $\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}$ имеет поддержку $\sup\left(\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}\right)$, если $\left(\sup\left(\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}\right) * P\right)$ записей БДДИ содержат множество событий $x_{l\tau_l g_{l\tau_l}} \cup f^{-1}\left(\Omega_{\tau_l}^{np}(s_l)\right)$.

Определение 8.

Ассоциативное правило $\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}$ справедливо с достоверностью $\mathrm{conf}\left(\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}\right)$, если $\mathrm{conf}\left(\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}\right) * \sup\left(f^{-1}\left(\Omega_{\tau_l}^{np}(s_l)\right)\right) * P$ записей, содержащих события $f^{-1}\left(\Omega_{\tau_l}^{np}(s_l)\right)$, также содержат множество событий $x_{l\tau_l g_{l\tau_l}}$,

$$\mathrm{conf}\left(\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}\right) = \sup\left(\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}\right) / \sup\left(f^{-1}\left(\Omega_{\tau_l}^{np}(s_l)\right)\right).$$

В общем случае процесс поиска ассоциативных правил состоит из трех основных этапов:

1. Поиск в исходном наборе данных часто встречающихся наборов элементов (ЧВНЭ), поддержка которых больше или равна заданному значению.

2. Генерация ассоциативных правил на основе найденных на предыдущем этапе ЧВНЭ. Если $\{\alpha,\beta,\chi,\delta\}$ - ЧВНЭ элементов, то на его основе можно построить правила $X \Rightarrow Y$, такие, что $X \cup Y = \{\alpha,\beta,\chi,\delta\}$. Поддержка правила $X \Rightarrow Y$ будет равна поддержке ЧВНЭ $\{\alpha,\beta,\chi,\delta\}$. Достоверность сгенерированного правила равна $\mathrm{conf}(X \Rightarrow Y) = \sup(X \Rightarrow Y) / \sup(X)$. Правило добавляется к результирующему набору, при условии, что его достоверность больше заданной.

3. Усечение списка найденных ассоциативных правил. Полученный результирующий набор правил анализируется человеком-экспертом, который удаляет правила, которые, по его мнению, не предоставляют достаточной практической ценности.

Выше описанные этапы позволяют решить классическую задачу поиска ассоциативных правил, которая состоит в поиске всех правил, отвечающих заданным пороговым значениям поддержки и достоверности. В нашем же случае необходимо выявить правила вида $\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}$, левая часть которых изначально известна, и задача сводится к поиску всех возможных вариантов правой части, при которых достоверность всего правила будет удовлетворять заданному порогу.

Следует учитывать тот факт, что поддержка всего правила $\sup\left(\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}\right)$ не должна выступать в роли ограничения, поскольку, задав ее пороговое значение, мы можем исключить из рассмотрения события, которые объясняют причины неисправного состояния службы. Связано это с тем, что более 90% процентов времени [8] служба находится в работоспособном состоянии. Следовательно, 90% данных, полученных в процессе мониторинга функционирования службы корпоративной IP-сети, содержат значения параметров, которые соответствуют работоспособному $\Omega^p(s_l)$ состоянию службы s_l и всего лишь 10% – значения, соответствующие множеству $\Omega^{np}(s_l)$ неработоспособных состояний службы s_l. Таким образом, задавая в качестве порогового достаточно низкое значение, равное 0.1, мы заведомо можем отсечь все события, которые описывают неработоспособные состояния из $\Omega^{np}(s_l)$.

На практике для решения задачи поиска ассоциативных правил в БДДИ корпоративной IP-сети важна достоверность правила $\mathrm{conf}\left(\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}\right)$, которая показывает, какова вероятность появления множества событий $x_{l\tau_l g_{l\tau_l}}$ при возникновении неработоспособного состояния $\Omega_{\tau_l}^{np}(s_l)$.

Исходя из вышеизложенного, процесс поиска ассоциативных правил вида $\Omega_{\tau_l}^{np}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}$ в БДДИ корпоративной IP-сети может быть описан ниже перечисленными этапами:

1. Формирование для каждого неработоспособного состояния службы $\Omega_{\tau_l}^{np}(s_l)$ набора $\Psi\left(\Omega_{\tau_l}^{np}(s_l)\right)$ записей БДДИ, которые содержат события $f^{-1}\left(\Omega_{\tau_l}^{np}(s_l)\right)$

2. Поиск в наборе $\Psi\left(\Omega_{\tau_l}^{np}(s_l)\right)$ множеств событий $x_{l\tau_l g_{l\tau_l}}$, которые удовлетворяют следующему требованию:

$$\sup\left(\Omega_{\tau_l}^{np}(s_l) \cup x_{l\tau_l g_{l\tau_l}}\right) \geq \mathrm{conf}^0 * \sup\left(f^{-1}\left(\Omega_{\tau_l}^{np}(s_l)\right)\right), \qquad (2)$$

где $conf^0$ - заданное пороговое значение достоверности для искомых ассоциативных правил.

Правая часть неравенства (2) представляет собой константу, обозначим ее как $\sigma = conf^0 * \sup\left(f^{-1}\left(\Omega^{np}_{\tau_l}(s_l)\right)\right)$, тогда (2) примет вид:

$$\sup\left(\Omega^{np}_{\tau_l}(s_l) \cup x_{l\tau_l g_{l\tau_l}}\right) \geq \sigma, \quad (3)$$

Поскольку мы работаем с набором $\Psi\left(\Omega^{np}_{\tau_l}(s_l)\right)$, где каждая запись заведомо содержит события $f^{-1}\left(\Omega^{np}_{\tau_l}(s_l)\right)$, тогда неравенство (3) можно записать в форме:

$$\sup_{\Psi\left(\Omega^{np}_{\tau_l}(s_l)\right)}\left(x_{l\tau_l g_{l\tau_l}}\right) \geq \sigma, \quad (4)$$

где $\sup_{\Psi\left(\Omega^{np}_{\tau_l}(s_l)\right)}\left(x_{l\tau_l g_{l\tau_l}}\right)$ - поддержка множества событий $x_{l\tau_l g_{l\tau_l}}$ в наборе записей $\Psi\left(\Omega^{np}_{\tau_l}(s_l)\right)$. Следовательно, поиск ассоциативных правил $\Omega^{np}_{\tau_l}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}$, достоверность которых выше или равна пороговому значению $conf^0$, сводится к поиску ЧВНЭ с поддержкой не менее σ в наборе записей $\Psi\left(\Omega^{np}_{\tau_l}(s_l)\right)$ БДДИ.

3. Удаление человеком-экспертом из результирующего списка ассоциативных правил, которые не предоставляют интереса при поиске причин неисправного состояния.

На сегодняшний день исследователи в области Data Mining [9] выделяют три типа методов поиска ЧВНЭ элементов в транзакционных наборах данных [10,11]:

1. всех ЧВНЭ;
2. максимальных ЧВНЭ;
3. замкнутых ЧВНЭ.

Пусть в 50% случаев служба s_l находится в неработоспособном состоянии $\Omega^{np}_2(s_l)$ по причинам, которые соответствуют варианту №2 (таблица 2), в 25% по причинам, которые соответствуют варианту №1 или №3 (таблица 2). Пусть для трех рассматриваемых вариантов состояние ЗДП описывается множествами событий $\{e_{212},e_{322},e_{422},e_{512}\}$, $\{e_{332},e_{412}\}$, $\{e_{212},e_{422}\}$.

Примем в качестве порогового значения достоверности искомых ассоциативных зависимостей значение, равное 0,15. Для заданного порогового значения в наборе записей $\Psi\left(\Omega^{np}_2(s_l)\right)$ проведем поиск частовстречающихся множеств с помощью методов трех вышеуказанных типов. Результаты поиска представлены в таблице 3.

Таблица 3

Результаты работы методов поиска ЧВНЭ

Тип метода	Состав результирующего множества X_{12}	Поддержка
1	$\{e_{212}\}$, $\{e_{322}\}$, $\{e_{422}\}$, $\{e_{512}\}$, $\{e_{212}, e_{322}\}$, $\{e_{212},e_{422}\}$, $\{e_{212}, e_{512}\}$, $\{e_{322}, e_{422}\}$, $\{e_{322}, e_{512}\}$, $\{e_{422}, e_{512}\}$, $\{e_{212}, e_{322}, e_{422}\}$, $\{e_{212}, e_{322}, e_{512}\}$, $\{e_{212}, e_{422}, e_{512}\}$, $\{e_{322}, e_{422}, e_{512}\}$, $\{e_{212},e_{322},e_{422},e_{512}\}$	0,25
	$\{e_{212}\}$, $\{e_{422}\}$, $\{e_{212},e_{422}\}$, $\{e_{322}\}$, $\{e_{412}\}$, $\{e_{332},e_{412}\}$	0,5
2	$\{e_{212},e_{322},e_{422},e_{512}\}$	0,25
	$\{e_{332},e_{412}\}$	0,5
3	$\{e_{212},e_{322},e_{422},e_{512}\}$	0,25
	$\{e_{332},e_{412}\}$, $\{e_{212},e_{422}\}$	0,5

Методы поиска всех ЧВНЭ находят необходимые множества $\{e_{212},e_{322},e_{422},e_{512}\}$, $\{e_{332},e_{412}\}$, $\{e_{212},e_{422}\}$, однако результат содержит также избыточные подмножества, которые могут быть получены простым перебором элементов множества $\{e_{212},e_{322},e_{422},e_{512}\}$. Метод поиска максимальных ЧВНЭ решает проблему избыточных подмножеств, но при его использовании теряется информация о действительном значении поддержки набора $\{e_{212},e_{422}\}$.

Необходимый результат возвращают методы поиска замкнутых ЧВНЭ, которые ищут максимально возможные множества с уникальным значением поддержки. То есть в случае их использования в результирующем списке для любого множества не существует надмножества с таким же значением поддержки.

Принимая во внимание выкладки данного раздела, метод поиска закономерностей между значениями собственных параметров из множества $B(s_l)$ и значениями неопределенных параметров из множества $\left(B = \bigcup_i^I B_i\right) / B(s_l)$ может быть описан следующим образом:

1. Процесс мониторинга состояния корпоративной IP-сети организуем в соответствии с требованиями модели (1).

2. Используем для идентификации состояния службы s_l методику, представленную ранее в статье.

3. Для неисправного состояния $\Omega^{np}_{\tau_l}(s_l)$ формируем набор $\Psi\left(\Omega^{np}_{\tau_l}(s_l)\right)$ записей, которые содержат события $f^{-1}\left(\Omega^{np}_{\tau_l}(s_l)\right)$.

4. Задаем пороговое значение $conf^0$ достоверности искомых ассоциативных правил.

5. Используем метод поиска замкнутых ЧВНЭ, входными данными для которого являются набор записей $\Psi\left(\Omega^{np}_{\tau_l}(s_l)\right)$ и пороговое значение поддержки $\sup^0 = conf^0 * \frac{\left|\Psi\left(\Omega^{np}_{\tau_l}(s_l)\right)\right|}{|\Psi|}$.

6. Полученные на этапе 5 множества, используем для формирования множества $X_{l\tau}$ и ассоциативных правил вида $\Omega^{np}_{\tau_l}(s_l) \Rightarrow x_{l\tau_l g_{l\tau_l}}$.

7. Путем проведения экспертной оценки усекаем результирующий список ассоциативных правил и удаляем те из них, которые не предоставляют интереса при поиске причин неисправного состояния $\Omega^{np}_{\tau_l}(s_l)$.

8. Этапы 3-7 выполняем для всего множества неисправных состояний $\Omega^{np}(s_l)$ службы s_l.

9. Этапы 2-8 проделываем для всех служб IP-сети.

5. Выводы

Научная новизна работы состоит в применении методов DataMinig для поиска закономерностей между значениями диагностических параметров служб корпоративной IP-сети.

Выявленные закономерности в дальнейшем могут быть использованы для получения информации о существующих скрытых зависимостях в работе аппаратных и программных компонентов IP-сети.

Применение метода поиска закономерностей в БДДИ корпоративной IP-сети дает возможность автоматизировать процесс создания и поддержки в актуальном состоянии базы знаний, которая может быть включена в состав системы диагностики или системы поддержки принятия решения для обслуживающего персонала корпоративной IP-сети.

В работе также представлена методика определения состояния службы корпоративной IP-сети и введена классификация диагностических параметров.

В качестве дальнейшего направления исследования рассматривается разработка метода, который бы позволял, используя обнаруженные в результате анализа БДДИ закономерности, искать наборы событий, которые с заданной вероятностью идентифицируют причины неисправностей служб корпоративной IP-сети.

Литература

1. Thottan, M. Anomaly detection in IP networks [Текст] / M.Thottan, C. Ji // IEE Transaction on signal processing. – 2003. – Vol.21, №8. – P.2191-2204.
2. 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.
3. Klemettinen, M. Interactive exploration of interesting patterns in the Telecommunication network alarm sequence analyzer TASA [Текст] / M. Klemettinen, H. Mannila, H. Toivonen // Information and Software Technology. – 1999. – Vol.41(9). – P.557-567.
4. Ndousse, T. D. Computational intelligence for distributed fault management in networks using fuzzy cognitive maps [Текст] / T. D. Ndousse and T. Okuda // in Proc.IEEE ICC. – Dallas. – TX, Jun. 1996. – P.1558–1562.
5. 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: www/ URL: http://diss.rsl.ru/diss/05/0616/050616043.pdf.
6. Сашин, С.В. Диагностика сети абонентского доступа с использованием информационных технологий [Электронный ресурс] : дис. ... канд. техн. наук / С.В.Сашин. – М.: РГБ, 2005 (Из фондов Российской Государственной библиотеки). – Режим доступа: www/ URL: http://diss.rsl.ru/diss/05/0464/050464041.pdf
7. ГОСТ 20911-89 Техническая диагностика Термины и определения [Текст]. – Введ. 1989-09-26.
8. Олифер, В.Г. Компьютерные сети. Принципы, технологии, протоколы [Текст] : Учебник для вузов. 3-е изд. / В. Г. Олифер, Н. А. Олифер. – СПб.: Питер, 2006. – 958с.
9. Fayyad U. From data mining to knowledge discovery in databases [Текст] / U.Fayyad, G.Piatetsky-shapiro, P.Smyth // AI Magazine. – 1996. – Vol.17. – P.37-54.
10. Goethals, Bart. Advances in Frequent Itemset Mining Implementations: Report on FIMI'03 [Электронный ресурс] / Bart Goethals , M.J. Zaki. – – CiteSeer. Scientific Literature Digital Library and Search Engine. – Режим доступа: www/ URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.136&rep=rep1&type=pd.
11. Ceglar, A. Association mining [Электронный ресурс] / John F. Roddick, A. Ceglar // ACM Computing Surveys (CSUR) – 2006. – Vol. 38(2), – Issue 2. – Article No. 5. – Режим доступа: www/ http://doi.acm.org/10.1145/1132956.1132958 – 31.03.2009. – Загл. с экрана.