

КРИТЕРІЇ ВІДБОРУ ДАНИХ ДЛЯ СТАТИСТИЧНОГО АНАЛІЗУ ПАРАМЕТРІВ ЯКОСТІ ТЕЛЕКОМУНІКАЦІЙНИХ МЕРЕЖ

В роботі розглядаються питання оперативної діагностики особливостей телекомунікаційних мереж, що важливі для підвищення якості передачі інформації. Проаналізовано параметри мережі для вирішення завдань мережевого адміністрування, моніторингу мережевого устаткування, виявлення аномальної поведінки чи збоїв в роботі системи. Розглядаються способи отримання статистичних оцінок параметрів мережі, засновані на використанні методів регресійно-кореляційного аналізу. Визначено критерії відбору параметрів регресійно-кореляційного аналізу.

Ключові слова: мультисервісна мережа, регресійно-кореляційний аналіз, множинна регресія.

N. Yakymchuk, L. Markina. The criteria of data selection in statistical analysis of telecommunication networks parameters.

The questions of operative diagnostics of features in telecommunication networks are studied, which are being important for improving the quality of information transfer. To solve the tasks of network administration, monitoring of network equipment, and reveal system abnormal behavior or network disturbance the network state is analyzed. The ways of receipt of statistical estimations of network parameters, based on the regression-correlation analysis methods are considered. The criteria of parameters selection for regression-correlation analysis are considered.

Keywords: multiservice network, regression-correlation analysis, multiple regressions.

Н.Н. Якимчук, Л.Н. Маркіна. Критерии отбора данных для статистического анализа параметров качества телекоммуникационных сетей

Рассматривается вопрос оперативной диагностики особенностей телекоммуникационных сетей, который является существенным для повышения качества передачи информации. Проанализированы параметры сети для решения заданий сетевого администрирования, мониторинга сетевого оборудования, выявления аномального поведения или сбоев в работе системы. Рассматриваются способы получения статистических оценок параметров сети, основанные на использовании методов регрессионно-корреляционного анализа. Определены критерии отбора параметров регрессионно-корреляционного анализа.

Ключевые слова: мультисервисная сеть, регрессионно-корреляционный анализ, множественная регрессия.

Постановка проблеми.

Питання оцінки якості зв'язку в телекомунікаційних системах і на сьогодні залишається складною багатокомпонентною задачею. Відомі різні методи вимірювання якості послуг зв'язку – радіовимірювання допомогою мобільних лабораторій, моніторинг якості роботи мережевого устаткування і сервісів за допомогою спеціальних програмно-апаратних комплексів операторів зв'язку, тощо.

При чому, слід розрізнити такі методи, як вимірювання і оцінка. Методи вимірювання в сфері телекомунікацій вимагають застосування дорогих спеціальних сертифікованих вимірювальних засобів і комплексів, а також серверів статистичної обробки отриманих даних по спеціальних стандартизованих методиках.

Методи оцінки дозволяють використовувати прості і навіть часто безкоштовні засоби, доступні будь-якому користувачеві. Цей інструментарій реалізує один метод – активне тестування з'єднання від абонента до абонента, "з кінця в кінець" (E2E – end-to-end). За своєю суттю це статистичний експеримент, в ході якого імітуються з тією або іншою мірою наближення дії абонента при використанні послуги зв'язку, здійснюється збір статистичних даних і подальша їх обробка з метою розрахунку ключових показників якості послуги (KQI – Key Quality Indicator). Тобто ми говоримо не про вимірювання, а про тестування і оцінку показників якості на основі тестувань. Зокрема використовуються абонентські пристрої з відповідним програмним забезпеченням, крім того оператор мережі застосовує засоби, які надає виробник устаткування

зв'язку (лічильники подій, сигнали тривоги, тощо), і зовнішні засоби контролю, такі як аналізатори протоколів і "трасувальники" з'єднань. Спільний аналіз всіх цих даних в єдиній інформаційній системі дозволяє досягати максимального ефекту, створення єдиної моделі передачі інформації.

Аналіз літературних джерел.

Задача підвищення якості передачі інформації пов'язана, перш за все, з процедурою оперативної діагностики мережі. Параметричне діагностування телекомунікаційних систем сьогодні розглядається спрощено, шляхом неперервного чи періодичного спостереження за параметрами трафіку з використанням аналізаторів протоколів, що пояснюється простотою, наочністю і оперативністю контролю [3,4]. В новому інфокомунікаційному середовищі кінцевою послугою, як правило, стає не послуга зв'язку, а контент, що часто передається в мережу сторонніми провайдерами, параметри якого часто оператору невідомі. Проте, поряд з інформаційними даними мережа передає значний об'єм сигнального трафіку, що містить в собі відомості практично про всі аспекти роботи мережі зв'язку, мережеві служби, запити абонентів і надані послуги. Інтелектуальна обробка цих даних дозволить виокремити зміни інтенсивності трафіку, наявність і характер збоїв, параметри інтенсивності використання окремих частин мережі, тощо.

Прийнято вважати, що якщо результати діагностування не виходять за встановлені межі, то мережа працездатна. Недоліками такого підходу є неможливість виявити проблеми передачі трафіку на ранніх стадіях їх виникнення, коли параметри ще не досягли граничних значень та недоступність інформації від віддалених вузлів системи, безпосередньо не зв'язаних з контрольованим.

Для розробки та застосування статистичних методів мережевого моніторингу та аналізу, перш за все, необхідно побудувати математичні моделі мережевого трафіку [7]. Очевидно, це є задачею математичної статистики [5,8].

На етапі моніторингу виконується процедура збору первинних даних про роботу мережі, статистика про кількість циркулюючих в мережі пакетів, про стан портів концентраторів, комутаторів і маршрутизаторів, тощо. Етап аналізу являє собою складніший і інтелектуальний процес осмислення зібраної в процесі моніторингу інформації, з використанням спеціальних інструментальних засобів і методик обробки часових рядів.

Нові можливості дає нам використання факторного аналізу сигнального трафіку, кореляційний аналіз його окремих складових дає можливість прогнозувати появу змін в системі та успішно вирішувати задачу мережевого адміністрування. Кореляційні зв'язки між факторами і параметрами відображають основну діагностичну інформацію по трафіку, що проходить через той чи інший інтерфейс мережевого пристрою.

Виклад основного матеріалу.

За тривалістю активні методи контролю поділяються на неперервні (моніторинг) і короточасні (навантажувальні). Навантажувальні вимірювання можуть виконуватися за розкладом або за запитом і, частіше всього, використовуються для виявлення причин несправностей, які виявлені в результаті моніторингу або суб'єктивної оцінки користувача про низький рівень обслуговування. Також вони можуть проводитися перед введенням об'єкту в експлуатацію, після проведення планово-профілактичних робіт, або перед відповідальними сеансами зв'язку (наприклад, відеоконференція).

Моніторинг проводиться на постійній основі стаціонарними засобами вимірювань і складається з циклу вимірювальних процедур, Тривалість циклу зазвичай складає від 1 хвилини до 1 години. Чим коротший цикл вимірювань, тим вища оперативність реагування на події погіршення якості послуг. Збільшення тривалості циклу підвищує точність вимірювання.

В результаті моніторингу мережі отримуємо значення наступних показників якості:

- пропускна спроможність мережі - інтегральний показник, що характеризує об'єм інформації, що пропускається,
- реакція на характеристики профілю трафіку - для визначення цієї характеристики мережа моделюється як чорний ящик, розглядається реакція на зміну навантаження на мережу,
- кількість втрачених пакетів - для ТСП мережі 1-5% втрачених пакетів, згідно з експертними оцінками, знаходиться в межах норми, 40% втрачених пакетів - граничне значення, при якому мережа практично не працює,
- час доставки - вимірюють часом подвійного ходу (у прямому і зворотному напрямі), цей показник фіксується з використанням програми PING,

- нерівномірність доставки пакетів - ця характеристика впливає на роботу окремих додатків, наприклад, передачі аудіопотоку чи відеоконференції або в пакетній телефонії.

Технічні показники функціонування мережі, як правило, представляються таблицями статистичних даних:

$$\begin{pmatrix} y(1) & y(2) & \dots & y(i) & \dots & y(N) \\ x_1(1) & x_1(2) & \dots & x_1(i) & \dots & x_1(N) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_j(1) & x_j(2) & \dots & x_j(i) & \dots & x_j(N) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_k(1) & x_k(2) & \dots & x_k(i) & \dots & x_k(N) \end{pmatrix}.$$

Статистичні дані представляють собою вибірку деякої реалізації значень випадкових величин:

- i -а реалізація чисельного значення результату y_i , $i=1,2,\dots,N$;

- j -а реалізація чисельного значення j -го фактора x_j , $j=1,2,\dots,N$.

Між випадковою величиною результату і випадковою величиною існує кореляційна залежність. Застосувавши процедури множинної регресії можна визначити вид зв'язку між кількома незалежними змінними (факторами - рядки матриці спостережень \mathbf{X}) і залежною змінною (елементи вектора відгуків $\bar{\mathbf{y}}$). [6]. У загальному випадку, процедура побудови множинної регресії полягає в оцінюванні параметрів лінійного рівняння.

$$E \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} E[y_1] \\ E[y_2] \\ E[y_3] \\ \vdots \\ E[y_m] \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ 1 & x_{21} & x_{22} & \dots & x_{2n} \\ 1 & x_{31} & x_{32} & \dots & x_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} \cdot \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix},$$

або те ж саме в компактному вигляді:

$$E[\bar{\mathbf{y}}] = \mathbf{X}\bar{\boldsymbol{\theta}}.$$

Регресійні коефіцієнти представляють незалежні вклади кожної незалежної змінної в передбачення залежної змінної.

На основі цих даних формується модель нормальної роботи мережі. Порівняння стану активного трафіку з параметрами отриманої моделі в будь-який момент часу дозволяє прогнозувати і оцінити ризики відхилень параметрів мережі і проаналізувати причини їх появи.

Обов'язковим етапом при регресійно-кореляційного аналізу є відбір значущих параметрів для формування остаточної регресійної моделі. При цьому зазвичай використовують два протилежних критерії: *по-перше*, щоб зробити рівняння ефективним для передбачення, ми повинні прагнути включити в модель по можливості більше незалежних змінних з тим, щоб можна було більш надійно визначити прогнозовані величини; *по-друге*, через витрати, пов'язані з отриманням великої кількості інформації і її подальшою перевіркою, необхідно прагнути, щоб рівняння включало якнайменше незалежних змінних.

Компромід між цими критеріями може бути досягнутий за рахунок вибору "найкращого" рівняння, що включає оптимальну кількість незалежних змінних. В роботі для пошуку "найкращого" рівняння регресії застосований кроковий метод (покрокова регресія), коли незалежні змінні (регресори) одна за одною включаються в підмножину згідно попередньо заданого критерію. У той же час деяка змінна може бути замінена іншою змінною, яка не входить в набір, або видалена з нього. Процедура визначення числа регресорів називається правилом зупинки – включення або виключення змінної X_j при визначенні F-критерію, що служить статистикою для перевірки того, що змінна X_j значимо покращує чи не покращує якість передбачення [8].

Крім того, процедура відбору параметрів моделі є експертною задачею і повинна враховувати ряд особливостей, пов'язаних із нелінійним темпом розвитку мереж, значним

збільшенням їх масштабів за короткий час, зростанням складності структури, кількості типів сервісів і навантажень на кожен з елементів. Результати оцінки параметрів QOS залежать від наступних чинників:

- Виду послуги мережі зв'язку - для різних послуг ключові вимірювані параметри якості різні (наприклад, для широкосмугового доступу – важлива гарантована смуга пропускання, а окремі втрати або затримки трафіку не критичні; для IP-телефонії смуга пропускання мінімальна, проте критичні відсутність втрат і час доставки трафіку).

- Інтенсивності існуючого трафіку в кожному сегменті мережі. Наприклад, в години інтенсивної роботи мережі, тобто із збільшенням числа з'єднань, в сегментах мережі зростають параметри затримки передачі пакетів, втрати пакетів. Це приводить до погіршення якості послуг зв'язку. І навпаки, під час відсутності інтенсивної роботи послуга може бути якіснішою.

- Оскільки вхідний і вихідний трафіки маршрутизації можуть відрізнятися (традиційно вхідний трафік набагато більший за вихідний), всі вимірювання мають бути односторонніми. Користувачі або системні служби можуть з'єднати статистику двох напрямів вже на етапі аналізу, щоб оцінити роботу в цілому.

- Основна тенденція останнього часу – зсув фокусу з вимірювання об'єктивних характеристик мережі оператора (QOS) на оцінку суб'єктивного сприйняття якості сервісу кінцевим користувачем (QOE), що пов'язано з розширенням видів послуг, що надаються операторами мережі. Наприклад, якщо десять років тому при моніторингу послуги IPTV використовувалися виключно транспортні характеристики (втрати пакетів, затримки, наявність/відсутність потоку) QOS, то в даний час важко собі уявити повноцінний моніторинг послуги без аналізатора каналів (розсіпання або тремтіння картинки, завищений або занижений рівень звуку відносно інших каналів і так далі).

- Необхідно враховувати і відсутність норми на чутливість приймальних пристроїв абонентського устаткування, тому оцінки, отримані різними екземплярами однотипних абонентських пристроїв, можуть істотно відрізнятися.

- Крім того зауважимо відсутність еталонної бази для деяких показників якості. Наприклад, оцінка якості передачі мови і відео зображення, розбірливість мови, що вимірюється в балах MOS (Mean Opinion Score), може бути неадекватно відображена споживачем послуги.

Таким чином показником якості роботи не може бути просто величина, наприклад, середньої швидкості, це має бути інтегральний показник. Крім того, поряд з результатами вимірювань параметрів слід приводити умови, при яких вони були здійснені. Це підвищить об'єктивність і достовірність результатів, дозволить порівнювати отримані моделі з оперативними даними мережі в кожному режимі роботи.

Врахування таких особливостей полегшить спектральний аналіз даних, обчислення тренду та оцінку частоти появи викидів завантаження каналу, дозволить виокремити значимі події, визначити вагу впливу кожного параметра на результативний чинник.

Результатами діагностики мережі є визначення нормального профілю роботи мережі і виявлення і передбачення відхилень від нього для відповідної зміни конфігурації системи. Для прикладу: збільшення кількості викидів – короткочасне підвищення завантаженості каналу, сигналізує про появу перешкод, шумів на лінії, що вимагають повторної передачі пошкодженої інформації, в комп'ютерних мережах сигналізує про появу вірусів, паразитного трафіку. Зміна спектральної характеристики ряду пов'язана зі зміною коливання добового завантаження мережі, що вказує на зміну режимів роботи користувачів мережі. Зміна тренду показує ріст завантаженості мережі, що може призвести до перевищення пропускну здатності. Стрибокподібне збільшення середнього значення завантаженості сигналізує, що мережа працює в режимі граничного завантаження.

Висновки. Оперативно здійснювати ефективну оцінку роботи всіх компонентів такої складної системи як телекомунікаційні мережі з урахуванням їх взаємозв'язків і взаємовпливу можливо лише з використанням статистичних методів аналізу і обробки оперативної інформації. Неefективний розподіл ресурсів, недосконала організація взаємодії апаратури обмежують пропускну спроможність мережі, погіршують якість її функціонування. Регресійно-кореляційний аналіз зв'язків між факторами і параметрами роботи комунікаційного обладнання відображають основну діагностичну інформацію по трафіку, що проходить через той чи інший інтерфейс мережевого пристрою, дозволяє прогнозувати зміни інтенсивності робочого навантаження, передбачати і сигналізувати про небезпеку відхилення від нормального профілю роботи мережі.

Складністю проведення статистичного аналізу є необхідність врахування і обробки великої кількості інформації, втрата частини якої внаслідок технічних чи організаційних перешкод чи збоїв веде до появи зміщених статистичних оцінок, погіршення якості моделей. Наведені критерії відбору параметрів регресійно-кореляційного аналізу враховують множину обмежень, що пов'язані з імовірнісним характером та різноманітністю інформації в системах телекомунікацій.

ПЕРЕЛІК ПОСИЛАНЬ

1. Tanenbaum A.S. Computer networks, 5th ed. / Andrew S. Tanenbaum, David J. Wetherall. – Prentice Hall, Cloth, 2011. – 960 p.
2. Stallings W. Foundations of modern networking: SDN, NFV, QoE, IoT, and Cloud. – Pearson Education, Inc., Old Tappan, New Jersey, 2016. – 538 p.
3. Язловецкий Я.С. Средства мониторинга и анализа сетевого трафика, используемые при измерении параметров QoS // Стандартизация и метрология. – 2015. - №5. – С. 31 – 37.
4. Лесная Н.Н. Сравнительный анализ методов оценки характеристик интеллектуальной сети // Наукові записки Українського науково-дослідного інституту зв'язку. – 2009. – №2(10). – С. 97-102.
5. Афифи А. Статистический анализ: Подход с использованием ЭВМ / А. Афифи, С. Эйзен. – Москва: Мир, 1982. – 488 с.
6. Мостеллер Ф. Анализ данных и регрессия: вып. 1 / Ф. Мостеллер, Дж. Тьюки. – Москва: Финансы и статистика, 1982. – 317 с.
7. Торошанко Я.І. Задачі моніторингу та аналізу параметрів телекомунікаційних мереж / Я. І. Торошанко, А.О. Булаковська, М. С. Височіненко, В. С. Шматко // Телекомунікаційні та інформаційні технології. – 2014. – №3. – С. 62-69.
8. Барабаш Ю.Л. Вопросы статистической теории распознавания / Ю. Л. Барабаш, Б. В. Варский, В. Т. Зиновьев, В. С. Кириченко, В. Ф. Сапегин. – Москва : Советское радио, 1967. – 400 с.