УДК 681.3.06+519.6(075.8)

Некряч О.В., аспірант (Тел. +380 (91) 114 35 34. Email: onekriach@gmail.com)
(Державний університет телекомунікацій, м.Київ)

АНАЛІЗ ВЗАЄМОЗАЛЕЖНОСТІ ПОДІЙ В СКЛАДНИХ ДИНАМІЧНИХ СИСТЕМАХ

Некряч О.В. Аналіз взаємозалежності подій в складних динамічних системах. Виконано аналіз основних питань які виникають при автоматизації процесу пошуку першопричини виникнення інцидентів. Наведено основні вимоги до процесу аналізу причини виникнення інциденту. Виходячи із сформульованих вимог, запропоновано структуру для системи виявлення першопричини інциденту. Запропоновано метод автоматичного виявлення залежностей між параметрами на основі підрахунку взаємної інформації. Згідно запропонованого методу виконано та наведено результати розрахунків.

Ключові слова: мережа передачі даних, складна система, динамічна структура, взаємозалежність подій, контрольовані параметри, взаємна інформація, першопричина інциденту

Некряч А. В. Анализ взаимозависимости событий в сложных динамических системах. Выполнен анализ основных вопросов, которые возникают при автоматизации процесса поиска первопричины возникновения инцидентов. Приведены основные требования к процессу анализа причины возникновения инцидента. Исходя из сформулированных требований, предложена структура для системы обнаружения первопричины инцидента. Предложен метод автоматического обнаружения зависимостей между параметрами на основе подсчета взаимной информации. Согласно предложенного метода выполнено и приведены результаты расчетов.

Ключевые слова: сеть передачи данных, сложная система, динамическая структура, взаимозависимость событий, контролируемые параметры, взаимная информация, первопричина инцидента

1. Вступ та постановка задачі

Інфокомунікаційні мережі сучасних великих підприємств можуть охоплювати майже всю країну. Вони являють собою складні, високоорганізовані структури з великою кількістю елементів, що входять до їх складу. Мережі включають в себе велику кількість маршрутизаторів, комутаторів, міжмережевих екранів, серверів та каналів зв'язку між ними різної пропускної спроможності.

Елементи мережі тісно пов'язані один з одним фізичними та логічними зв'язками і утворюють динамічну систему в якій постійно відбуваються зміни тих чи інших параметрів. Враховуючи наявність зв'язків між елементами мережі, зміна стану одного з них, з певною імовірністю, призводить до зміни параметрів інших елементів мережі.

В реальних умовах ці взаємозв'язки є складними та не завжди очевидними, тому у складі системи управління необхідно мати інструмент, який дозволить автоматично визначати наявність залежностей між контрольованими параметрами.

У великих мережах, у зв'язку з ростом числа клієнтів та змін у вимогах до мережевої інфраструктури, залежності між елементами стають все більш складними. Відповідно і локалізація кореневих причин тих чи інших процесів є складним завданням.

2. Вимоги до процесу аналізу виникнення інциденту

В ході аналізу необхідно зібрати стільки інформації про мережу, скільки можливо. У зв'язку з постійними змінами в мережі і систем, не можна повністю довіряти існуючій експлуатаційній документації, тому одним з перших компонентів інструменту аналізу першопричин має бути механізм відкриття структури мережі та компонентів, що до неї входять. При цьому розкриття структури потрібно виконувати порівнево починаючи з фізичної топології і закінчуючи логічними взаємозв'язками та схемою потоків даних [1].

Другий важливий компонент – зафіксувати несправність, тобто слід визначити несправну поведінка системи по одному з вузлів мережі або будь-якого компонента системи.

Комплексний аналіз повинен включати обробку проблемних звернень кінцевих користувачів з приводу несправностей у наданні сервісу, для того, щоб визначити основні проблеми з продуктивністю. При цьому здійснюється оцінка стану кожного компонента моделі протягом часу, коли надходили проблемні звернення, на основі відхилень від нормальної поведінки, яка спостерігалась при нормальному режимі роботи [2].

У великих і складних мережах засоби моніторингу генерують велику кількість даних які важко або часто просто не можливо аналізувати в ручному режимі. Велика кількість даних не завжди означає велику кількість інформації [2]. Корисну для системи управління інформації потрібно ще виділити та привести до певного формату. Тому для повноцінного моніторингу та управління необхідна система, що забезпечуватиме автоматичний аналіз подій.

При впровадженні систем аналізу взаємозалежностей подій виникає ряд питань:

- відсутність чіткої, повної і належним чином впорядкованої інформації про систему
- велика кількість процесів, що виконуються та відсутність їх опису
- виділення повного набору основних параметрів (метрик)
- величезні обсяги даних для обробки
- наявність численних складних взаємодій між компонентами системи
- наявність не очевидних взаємозалежностей між контрольованими параметрами системи.
- необхідність взаємодії із сторонніми системами моніторингу.

Виходячи із вищесказаного, система аналізу першопричини виникнення подій у складних системах буде мати структуру наведену на Рис.1.

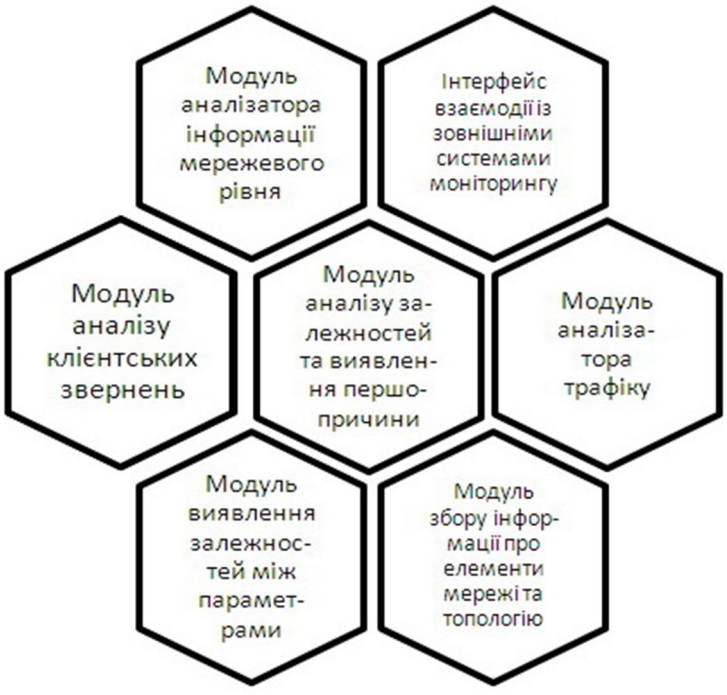

Рис. 1. Структура системи аналізу першопричини виникнення інциденту

3. Визначення наявності взаємозв'язку між контрольованими параметрами

Значні складнощі виникають в процесі виявлення взаємозалежностей між контрольованими параметрами. Залежності не завжди очевидні і якщо задавати їх тільки в ручному режимі, то деякі важливі кореляції можуть бути пропущені.

Для вирішення цієї задачі необхідно використовувати алгоритми автоматичного виявлення взаємозалежності між контрольованими параметрами.

Взаємозалежність між контрольованими параметрами може бути визначена на основі аналізу імовірнісних характеристик цих параметрів. Так, проаналізувавши наявність взаємної інформації між двома параметрами можна встановити наявність кореляційної взаємозалежності між ними.

Взаємна інформація визначається через *ентропію*. Поняття ентропія запозичене із термодинаміки, де є аналогічний за своєю формою вираз, що характеризує невизначеність стану фізичної системи [3].

У теорії інформації ентропія *H(X)* також характеризує невизначеність ситуації до передавання повідомлення, оскільки наперед невідомо, яке з повідомлень ансамблю джерела буде передано. Чим більша ентропія, тим сильніша невизначеність і тим більшу інформацію в середньому несе одне повідомлення джерела [2].

$$H\left(X\right)=-\sum_{i=1}^{m}P(x_i)logP(x_i)\,, \tag{1}$$

де $P(x_i)$ – імовірність події x_i.

Якщо ентропія характеризує невизначеність однієї випадкової величини, можна визначити умовну ентропію $H\left(X_i \mid Y_j\right)$, де ентропія випадкової величини (*X*) умовно пояснюється іншою змінною (*Y*).

Якщо дві випадкові величини *X* і *Y* якимось чином пов'язані одна з одною (інакше кажучи, існує кореляційна залежність між *X* і *Y*), то знання будь-якої з них, очевидно, зменшує невизначеність іншої. Невизначеність, що залишається, оцінюється умовною ентропією. Умовна ентропія *X* за умови відомого *Y* визначається як [5]:

$$H\left(X|Y\right)=\sum_{i=1}^{K}\sum_{j=1}^{N}P(X_i \mid Y_j)\log P\left(X_i|Y_j\right), \tag{2}$$

де $P\left(X_i \mid Y_j\right)$ – умовна ймовірність (імовірність *i*-го значення *X* за умови *Y*=Y_j;

Зменшення невизначеності однієї випадкової величини завдяки отриманню значень іншої змінної називається "взаємною інформацією". Тоді, враховуючи (1) і (2), міра залежності між двома змінними обчислюється таким чином:

$$I\left(X,\ \mathrm{Y}\right)=\ H\left(X\right)\ -\ H(\ \mathrm{X}\mid Y)=\sum_{x,y}^{k,l}P\left(X_k,Y_l\right)\log\frac{P\left(X_k,Y_l\right)}{P\left(X_k\right)P\left(Y_l\right)}\,. \tag{3}$$

Взаємна інформація завжди ненегативна і дорівнює нулю, тільки якщо *X* і *Y* є незалежними. Взаємна інформація максимальна і дорівнює безумовній ентропії, коли між *X* і *Y* є однозначна залежність [2].

Виходячи з наведеної вище інформації із врахуванням (3), можна описати базовий алгоритм автоматичного виявлення залежностей між контрольованими параметрами:

1) Визначення контрольованих параметрів (*X, Y,..., m*)
2) Отримання значень контрольованих параметрів (X_k, Y_l, де *k*, l є 0, 1,…, *n*)

3) Визначення імовірності кожного із значень ($P(X_k)$, $P(Y_l)$)
4) Попарне визначення умовної імовірності між значеннями ($P(X_k, Y_l)$)
5) Визначення ентропії кожного із параметрів ($H(X)$)
6) Попарне визначення умовної ентропії між значеннями ($H(\mathrm{X}|Y)$)
7) Визначення взаємної інформації між параметрами інтервалі ($I(X, \mathrm{Y})$)
8) Формування матриці взаємної інформації ($m \times n$)
9) Визначення комбінацій параметрів із найбільшим значенням взаємної інформації.
10) Побудова матриці залежних параметрів.

Кожне із значень береться для розрахунку в певному інтервалі ($\pm m$ де $m \in 0,1,...n$). Використання інтервалу дозволяє знехтувати короткочасними випадковими флуктуаціями контрольованого параметру.

4. Результати розрахунків та їх інтерпретація

Використовуючи даний алгоритм виконано аналіз чотирьох масивів даних. Два з них – значення утилізацію центрального процесора маршрутизатора (CPU, %) та значення затримки (Latency, мс) при ICMP-опитуванні цього маршрутизатора системою моніторингу:

CPU =[30,30,30,30,100,100,100,100,45,30,30,30]
Latency =[1,1,1,1,60,60,60,60,1,1,1,1]

Два інших – завідомо не залежні випадкові величини Random1 та Random2:

Random1=[0,1,1,0,0,0,0,1,0,1,0,1]
Random2=[1,1,1,0,0,1,0,0,1,1,0,1]

Табл. 1 містить результати попарного розрахунку взаємної інформації (MI – mutual information) між кожним із векторів контрольованих параметрів (матриця взаємної інформації).

Матриця взаємної інформації **Табл. 1**

MI	CPU	Latency	Random1	Random2
CPU	3.88	3.12	0.38	0.76
Latency	3.12	3.61	0.15	0.96
Random1	0.38	0.15	4.03	0.52
Random2	0.76	0.96	0.52	4.03

Як видно із Табл.1 найбільші значення взаємної інформації мають параметри CPU та Latency. Діагональ матриці взаємної інформації при цьому до уваги не береться.

Таким чином, матриця залежних параметрів (Табл. 2) матиме наступний вигляд:

Матриця залежних параметрів **Табл. 2**

Параметр 1	Параметр 2	MI
CPU	Latency	3.12

У багатьох випадках корисною є не тільки інформація про наявність залежності між параметрами а й про її відсутність. Так у Табл.3 наведено параметри між якими взаємозалежності не виявлено.

Матриця незалежних параметрів **Табл. 3**

MI	CPU	Latency	Random1	Random2
CPU	0	-	0.38	0.76
Latency	-	0	0.15	0.96
Random1	0.38	0.15	0	0.52
Random2	0.76	0.96	0.52	0

5. Висновки

Складність взаємозв'язків у сучасних мережах є динамічним показником. Їх поглиблення та зростання масштабу буде зростати і надалі оскільки розвиток мереж не може стояти на місці і в свою чергу корелюється із розвитком бізнес сегменту, промисловості та все більшої необхідності суспільства у постійних комунікаціях.

Наведений вище метод дозволяє автоматизувати процес виявлення взаємозалежностей між контрольованими параметрами телекомунікаційної мережі. Отримані результати розрахунків підтверджують його відповідність та можливість використання як окремого блоку аналізу даних у системі моніторингу та як складової частини системи виявлення першопричини інцидентів в мережевій інфраструктурі.

Література:

1. Hyong S. Kim. Root cause analysis in large and complex networks. University of Lisboan. – 2008. –66 p.

2. I. Paredes-Oliva (UPC), M. Molina (Cambridge), Automating Root-Cause Analysis of Network Anomalies using Frequent Itemset Mining. – 2010.

3. Robert M. Gray. Entropy and Information Theory. – First Edition. – Stanford University. –March 2013. – 31 p.

4. Беркман Л. Н., Кривуца В. Г., Стеклов В. К. Управління телекомунікаціями із застосуванням новітніх технологій. – К.: Техніка, 2007. – 384 с.

5. Thomas M. Cover, Joy A. Thomas. Elements Of Information Theory. – Second Edition. Canada. – 2006. – 774 p.

Дата надходження в редакцію: 2.04.2015 р. Рецензент: д.т.н., проф. Л. Н. Беркман