УДК 004.056 (043.2)

С.Г. Семенов[1], І.А. Березюк[2]

[1] *Національний технічний університет «Харківський політехнічний інститут», Харків*
[2] *Кіровоградський національний технічний університет, Кіровоград*

ВІДНОВЛЕННЯ СЕМАНТИКИ СЛУЖБОВОГО ПОВІДОМЛЕННЯ В ЗАДАЧІ СТЕГАНОАНАЛІЗУ

У статті запропоновано підхід щодо відновлення семантики службового повідомлення в задачі стеганоаналізу. Відмінною особливістю даного походу є врахування ключових службових полів, що використовуються телекомунікаційними протоколами. Представлений алгоритм виявлення семантики є загальним. Розроблено алгоритм виведення семантики. Визначено складність даного алгоритму, з урахуванням наявності даних алгоритму поширення помічених даних, і результатів обчислення комірок з відомою семантикою.

Ключеві слова: *стеганографія, стеганоаналіз, семантика службових повідомлень.*

Вступ

Постановка задачі та аналіз літератури. У сучасному суспільстві однією з найбільш важливих задач є задача інформаційного обміну. Розвиток телекомунікаційних технологій дозволяє забезпечувати різні користувальницькі послуги інформаційного обміну з певною якістю. При цьому технологічна підтримка обміну даними виконується за допомогою стандартних технологій і протоколів мережевого обміну (наприклад, TCP / IP).

Відомо, що основні службові поля, що використовуються для управління процесом передачі даних в телекомунікаційних технологіях, є текстовий набір спеціалізованих команд і семантичних кодів, розміщених певним встановленим чином. Найчастіше дані поля візуально недоступні звичайним користувачам і з точки зору міжкористувального інформаційного обміну не представляє значного інтересу.

Однак, як показали дослідження, даний вид службових повідомлень становить інтерес фахівців (легальних і злочинних користувачів) як контейнери для прихованої (стеганографічної) передачі конфіденційних даних.

Проведений аналіз літератури [2-6] показав, що в даний час існує декілька найбільш загальних методів вбудовування повідомлень в текстові дані (в тому числі текстові службові поля телекомунікаційних протоколів). Їх можна умовно розділити на три види: синтаксичні методи, методи, що генерують текст, подібний природному, семантичні методи. Найбільш докладно ці методи описані в [2 – 6].

Проведені дослідження показали, що одним з найбільш складних, з точки зору стеганоаналізу, є семантичні методи, які дозволяють приховано передавати дані використовуючи в якості контейнера, в тому числі, і службові поля. Частково причиною цього є незначні зміни статистичних даних використовуваних при стеганоаналізі. При цьому відновлення семантики службових повідомлень є складною і актуальною науковою і технічною задачею.

Основна частина

Як правило, більшість полів в повідомленні містять дані, специфічні для протоколу і зрозуміти сенс значень цих даних можна тільки знаючи завдання, що вирішує конкретний протокол. У той же час, існує ряд типів даних, які використовують багато протоколів, що дозволяє розробити загальні підходи до пошуку полів, що містять дані цих типів. До таких даних можна віднести, наприклад, *IP*-адреси, імена файлів, імена серверів в мережі, номери портів, часові мітки.

Пропонований в статті підхід до відновлення семантики заснований на методі виведення семантики. Відповідно до даного підходу вводиться поняття "джерело семантики", тобто місце в програмі, в якому семантика використовуваних даних точно відома. Такими джерелами можуть бути наступні.

– спеціальні інструкції, наприклад інструкція запиту часу *RDTSC*, архітектури *x86*, результат якої - число тактів, записується в певні регістри,

– виклики функцій стандартної бібліотеки, семантика параметрів яких відома. Наприклад, функції для роботи по мережі оперують *IP*-адресами, а функції роботи з файловою системою - іменами файлів,

– спеціальні конструкції в *CFG* програми, наприклад умовні переходи, що переривають виконання циклу за умовою. Якщо в циклі обробляється масив і значення кожного його елемента порівнюється з константою, то поле, що містить константу, є полем роздільником, а сама константа – кінцевим символом.

Формально, для випадку аналізу бінарних трас, термін джерело семантики можна ввести, як трійку <t, m, s>, де t - крок траси, а m - елемент пам'яті або регістр, s- тип семантики значення t. Так як в трасі, одно і той же джерело семантики може зустрічатися багаторазово (наприклад виклик одній і тій же функції), то для аналізу потрібне мати загальний статичний механізм опису джерел семантики.

Такий механізм можна висловити трійкою <F(t), M(t), S(t)>, де F – булева функція, яка для заданого

кроку повертає істину, якщо крок є джерелом семантики. M(t) – функція, що дозволяє по заданому кроці отримати елемент пам'яті або регістр, в якому міститься, значення, семантика якого відома. S(t) – функція, що дозволяє по заданому кроці отримати тип семантики. Алгоритм виведення семантики в загальному вигляді може бути представлений у вигляді схеми рис. 1.

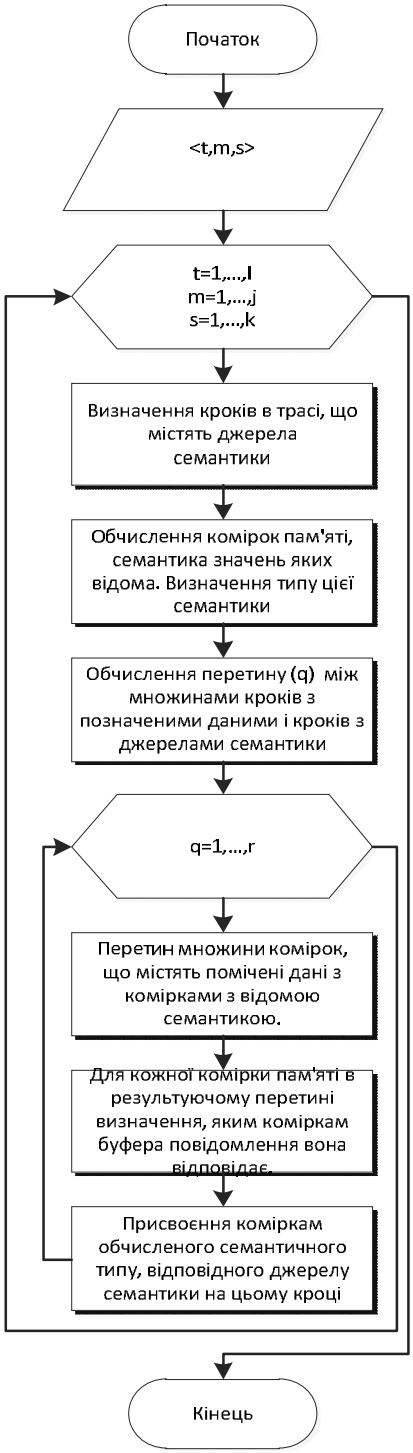

Рис. 1. Блок схема алгоритму виведення семантики

Складність алгоритму, з урахуванням наявності даних алгоритму поширення помічених даних, і результатів обчислення комірок з відомою семантикою можна оцінити як 0 (N), N – кількість кроків в трасі.

Зупинимося докладніше на описі джерел семантики. Список можливих типів джерел був приведений вище. Опишемо докладніше, як виглядають механізми обчислення комірок і типів семантики для кожного типу джерела.

Для спеціальних інструкцій, типу RDTSC осередками з семантичними значеннями будуть вхідні або вихідні операнди цих інструкцій. Саме для RDTSC такими осередками є пара регістрів EDX: Eax.

Для викликів функцій стандартної бібліотеки точками семантики в трасі будуть виклики цих функцій. Описи комірок з відомою семантикою будуть являти собою опис формальних параметрів, а механізм обчислення конкретних комірок для кожного виклику відповідає способу отримання фактичних параметрів виклику за формальними параметрами. Тип семантики визначається по семантиці відповідного формального параметра. Опишемо далі спеціальні конструкції в CFG, що дозволяють визначати деяких види семантики.

Значення поля довжини тим чи іншим способом задає розмір послідовності, якої це поле відповідає. Воно може містити як розмір послідовності в байтах, так і кількість елементів в послідовності. Крім того, можливо, що однією послідовності відповідає кілька полів довжини. Кількість ітерацій в послідовності, що оброблює всі пікселі обчислюється як добуток цих двох полів, тобто довжину послідовності пікселів визначають не значення цих полів окремо, а комбінація (добуток) значень цих двох полів.

У будь-якому випадку, значення поля довжини має використовуватися в умовах виходу з циклу тобто в операції порівняння на кожній ітерації циклу.

Ще одним артефактом в досліджуваному пошуку є полів роздільників. Проведені дослідження показали, що полі роздільник – це поле, що слідує відразу після деякої послідовності, а його значення інтерпретується оброблювачем, як термінальний символ, що завершує послідовність. За визначенням його значення – це деяка константа або набір констант (як, наприклад, в разі HTTP), значення яких відомо оброблювачу. Для того щоб визначити кінець послідовності, обробник порівнює кожен елемент послідовності з термінальним символом (константою).

У процесі аналізу ключових полів в повідомленні, розбірник перевіряє значення таких полів, порівнюючи їх з цим набором значень і, в залежності від результату порівнянь, здійснює подальший розбір. Таким чином, значення цього поля визначає конкретний вид і інтерпретацію деякої частини повідомлення. Пошук ключових полів визначається способом їх обробки – потрібно знайти поля, значення яких порівнюється з деяким набором констант, причому одне з порівнянь істинно, і за результатами порівняння здійснюється умовний перехід в обробнику. Пошук міток полів аналогічний пошуку ключових значень, за винятком того, що в якості шаблонів пошуку використовуються такі:

BT field, bitNum; ConditionalJump target;
та *And field, const; ConditionalJump target.*

Результатом роботи не є набори константних значень, а межі окремих груп міток в рамках поля, які визначаються сукупністю застосовуваних до полю масок.

Проведені дослідження [1, 5] показали, що в багатьох протоколах і форматах фалів присутні поля, значення яких не можуть приймати будь-які значення, які не залежать від значень інших полів. Це означає, що деяке поле містить значення, яке є функцією від значень деякого набору інших полів. Найбільш часто використовуються поля що зберігають контрольні суми, обчислені по набору значень деяких інших (або полів) байт повідомлення. У цю ж групу можна віднести поля дублери, які зберігають однакові значення. Такі поля можуть виникати в процесі розвитку деякого протоколу або формату для забезпечення зворотної сумісності. Прикладом поля контрольної суми є поле, що зберігає значення *CRC32* в файловому форматі *ZIP*, а поля дублери зустрічаються в деяких заголовках різних версій формату *BMP*. Інформація про пов'язані групи полів є важливою, так як цей зв'язок накладає значні обмеження на можливі значення полів всередині цих груп. При обробці пов'язаних полів розбірник, як правило, виконує перевірку коректності повідомлення. В ході цієї перевірки він обчислює значення функції по набору полів і порівнює вийшло значення зі значенням пов'язаного поля. У разі полів дублерів функція є константою, що дорівнює значенню одного з полів дублерів.

Алгоритм пошуку пов'язаних полів визначається способом їх обробки – потрібно знайти поля повідомлення, значення яких порівнюється зі значенням, яке є функцією від значень деякого набору полів. Тобто шаблон пошуку виглядає таким чином:

Compare fieldl, func (f ield2, f ield3,..., f ieldN)
ConditionalJump target

Висновки

Таким чином, в статті представлений похід до відновлення семантики службового повідомлення в завданні стеганоаналізу. Відмінною особливістю даного походу є врахування ключових службових полів, що використовуються телекомунікаційними протоколами. У всіх алгоритмах, де передбачається робота з можливими значеннями, для гарантії коректності одержуваних значень потрібно знати, що на тих позиціях, де проводиться порівняння, значення комірки, яка піддається порівнянні дорівнює значенню вихідного поля повідомлення. До таких алгоритмах відносяться пошук полів роздільників, ключових полів, полів прапорів і пов'язаних полів. Вимога рівності значень означає, що всі інструкції, через які значення вихідного поля потрапило в осередок, значення якої порівнюється з константою були інструкціями копіювання. Для задоволення цієї вимоги пропонується використовувати ознака незмінності значення.

Список літератури

1. Аветисян А.И. Восстановление структуры бинарных данных по трассам программ [Електронний ресурс]/ А.И. Аветисян, А.И. Гетьман // Труды Института системного программирования РАН, т. 22. – 2012 . – Режим доступу: http://cyberleninka.ru/article/n/vosstanovlenie-struktury- binarnyh-dannyh-po-trassam-programm.

2. Конахович Г.Ф. Компьютерная стеганография / Г.Ф. Конахович, А.Ю. Пузыренко Теория и практика. — К.: МК-Пресс, 2006. – 288 с.

3. Нечта И.В. Метод стегоанализа текстовых данных, основанный на использовании статистического анализа / И.В. Нечта // Вестник СибГУТИ. – 2011. – № 3. – С. 27-34.

4. Семенов С.Г. Протоколи захисту інформації у комп'ютерних системах та мережах / С.Г. Семенов, О.О. Кузнецов. – Х.: ХНУРЕ, 2009. – 184 с.

5. Winstein K. Tyrannosaurus lex 1999 [Електронний ресурс]. – Режим доступу: http://alumni.imsa.edu/~keithw/tlex.

6. Meng P., Huang L., Chen Z, Yang W., Li D. Linguistic Steganography Detection Based on Per-plexity / Електронний ресурс]: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=5089098&isnumber=5089035

Надійшла до редколегії 14.07.2016

Рецензент: д-р техн. наук, проф. О.О. Можаєв, Національний технічний університет «ХПІ», Харків.

ВОССТАНОВЛЕНИЕ СЕМАНТИКИ СЛУЖЕБНОГО СООБЩЕНИЕ В ЗАДАЧЕ СТЕГАНОАНАЛИЗА

С.Г. Семенов, И.А. Березюк

В статье предложен подход к восстановлению семантики служебного сообщения в задаче стеганоанализа. Отличительной особенностью данного похода является учет ключевых служебных полей, используемых телекоммуникационными протоколами. Представленый алгоритм выявления семантики является общим. Разработан алгоритм вывода семантики. Определена сложность данного алгоритма с учетом наличия данных алгоритма распространения помеченных данных и результатов вычисления ячеек с известной семантикой.

Ключевые слова: *стеганография, стеганоанализ, семантика служебных сообщений.*

RECOVERY SEMANTIC SERVICE ANNOUNCEMENT IN PROBLEMS STEHANOANALIZ

S.G. Semenov, I.A. Berezyuk

The paper proposes an approach to restore service message semantics stehanoanaliz the problem. A distinctive feature of this campaign is the consideration of the key fields of service used telecommunications protocols. Presented detection algorithm is common semantics. The algorithm output semantics. Determined complexity of the algorithm, taking into account the availability of data dissemination algorithm marked data and results of calculations cells with known semantics.

Keywords: *steganography, stehanoanaliz semantics service messages.*