Защита информации от утечки по скрытым каналам

Лабораторная работа №2. Поиск скрытого сообщения в потоке трафика.

Содержание

1. Теория

Целью данной лабораторной работы является изучение метода обнаружения сетевых скрытых каналов по времени, основанных на изменении длин межпакетных интервалов.

Для противодействия скрытым каналам по времени более предпочтительны методы, не влияющие на функционирование легитимных каналов связи, так как в случае скрытых каналов по времени такие методы приводят к более ощутимым последствиям для работы всей системы в целом, чем аналогичные методы для скрытых каналов по памяти. Такой возможной не превентивной мерой противодействия является обнаружение.

Задача обнаружения функционирующих скрытых каналов в защищаемой системе сводится либо к задаче поиска закономерностей в потоке трафика, либо к задаче сравнения тестируемой выборки с «эталонной». Для сравнения выборок применяются известные методы математической статистики (например, критерий Колмогорова-Смирнова и критерий Пирсона). Для поиска закономерностей в потоке трафика применяются специально разработанные методы. Отдельно выделяется область методов обнаружения на основе алгоритмов машинного обучения.

Стоит отметить, что чем проще схема передачи скрытой информации, чем проще данный скрытый канал обнаружить. Поэтому схемы скрытых каналов специально усложняют.

Для случая простейшего бинарного скрытого канала по времени на основе изменения длин межпакетных интервалов существует специальный метод обнаружения, рассмотренный ниже.

Пусть дан бинарный скрытый канал, основанный на изменении длин межпакетных интервалов. Будет выбрано два отрезка времени, один из которых будет кодироваться нулем, другой — единицей. Для передачи скрытой информации будут посылаться пакеты с задержкой, попадающей в один из двух данных интервалов. Таким образом, если построить гистограмму поведения трафика в сети для канала связи, в котором присутствует описанный скрытый канал, представленную на рисунке 1, где будет представлено распределение числа пакетов в зависимости от длин межпакетных интервалов, то на данной гистограмме будут видны два пика с максимальным числом переданных пакетов C_{max}, сосредоточенных возле выбранных на этапе кодирования временных интервалов. Среднее значение времени межпакетных интервалов μ, в свою очередь, будет находиться между двумя данными пиками. При этом количество пакетов C_μ в точке μ будет мало.

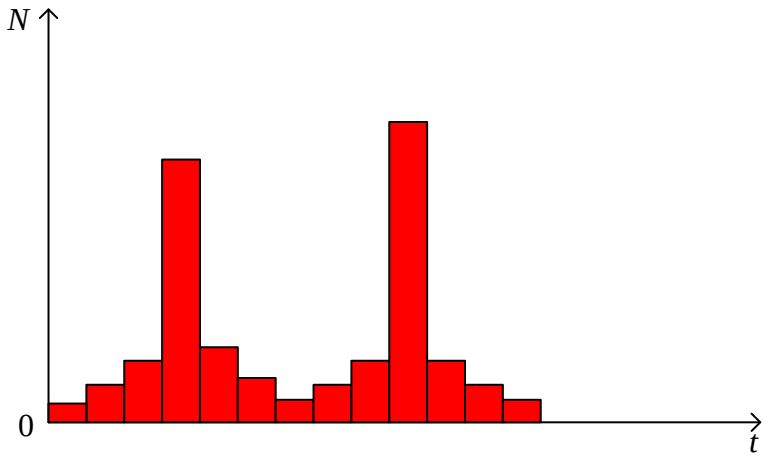

Рисунок 1 — Гистограмма поведения трафика в сети для канала связи, в котором присутствует бинарный скрытый канал

С другой стороны, если построить гистограмму, изображенную на рисунке 2, для канала связи без присутствия скрытого канала, то она будет заметно отличаться. Присутствует один пик, находящийся примерно по центру, в котором количество пакетов C_μ максимально по отношению к другим, то есть $C_\mu = C_{max}$.

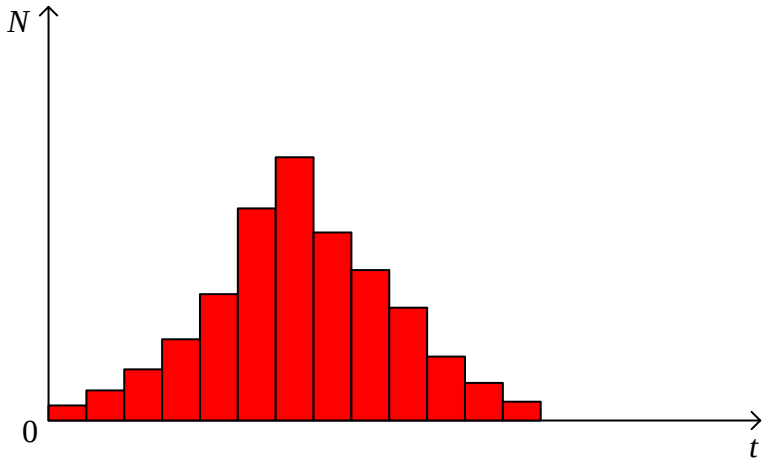

Рисунок 2 — Гистограмма поведения трафика в сети для канала связи, в котором отсутствует бинарный скрытый канал

Вводится отношение $\frac{C_\mu}{C_{max}}$. Легко заметить, что при $\lim\limits_{N\to\infty} \frac{C_\mu(N)}{C_{max}(N)} = 1$, где N — количество переданных пакетов, можно утверждать, что с высокой долей вероятности скрытого канала в системе нет, и, наоборот, при $\lim\limits_{N\to\infty} \frac{C_\mu(N)}{C_{max}(N)} \ll 1$ скорее всего, скрытый канал присутствует в системе.

Таким образом, конечная формула вероятности P наличия скрытого канала в системе имеет вид:

$$P = 1 - \lim_{N \to \infty} \frac{C_\mu(N)}{C_{max}(N)}. \quad (1)$$

Стоит отметить, что вероятность обнаружения построенного скрытого канала данным методом понижается, если выбирать временные интервалы для кодирования таким образом, чтобы они были расположены близко друг к другу, тем самым приближая значение $\frac{C_\mu}{C_{max}}$ к единице, как показано на рисунке 3. Однако, это ведет к увеличению уровня шума.

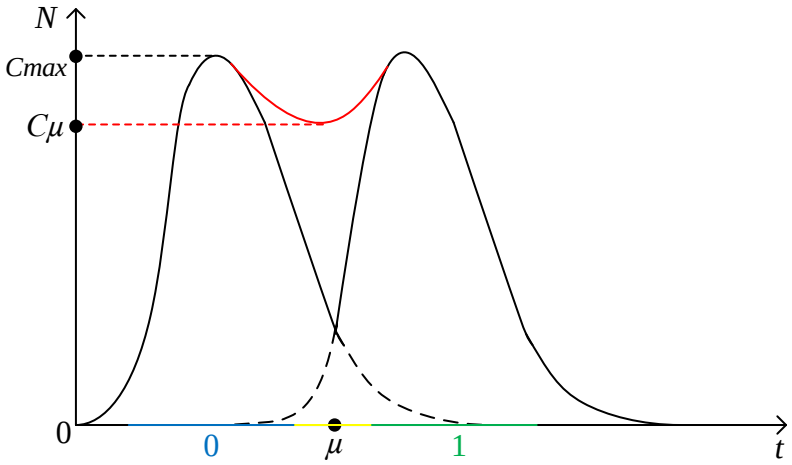

Рисунок 3 — Приближение значения $\frac{C_\mu}{C_{max}}$ к единице при неудачном выборе параметров кодирования

2. Задание

1. В качестве исходных данных дается дамп трафика. Известно, что при взаимодействии двух сторон было передано скрытое сообщение. Список задач:

- построить гистограмму отсортированных по возрастанию длин межпакетных интервалов;

Удобным средством для визуализации результатов анализа данных в Python'е является библиотека matplotlib.

- определить значение вероятности присутствия скрытого канала в системе согласно формуле (1).

2. По новой информации, скрытое сообщение передается, начиная с ~100 пакета и до последнего пакета в дампе трафика.

- построить гистограмму отсортированных по возрастанию длин межпакетных интервалов для трафика, в который включен только промежуток с передачей скрытого сообщения;

- определить значение вероятности присутствия скрытого канала в системе согласно формуле (1).

3. Необходимо понять, какое скрытое сообщение передавалось.

- разработать средство декодирования скрытого сообщения и определить, какое скрытое сообщение было передано.

4. Необходимо провести анализ применяемого в лабораторной работе метода обнаружения.

- сделать вывод о применимости метода обнаружения в случае малого объема переданной скрытно информации;

- предложить способ первоначальной обработки трафика в случае, если объем трафика большой, а количество переданной скрытно информации — малый;

- предложить способ расширения метода обнаружения на мультисимвольные скрытые каналы на основе изменения длин межпакетных интервалов.

Для защиты лабораторной работы необходимо предоставить отчет, содержащий в себе описание хода выполнения и результатов всех указанных в задании пунктов.

3. Варианты

Номер варианта определяется по формуле (N%12) + 1, где N — номер в списке группы. Номер варианта совпадает с номером дампа трафика.