DOI 10.36622/1682-7813.2024.27.4.009
УДК 004.056

МЕТОД СТАТИСТИЧЕСКОГО АНАЛИЗА ЛОКАЛЬНОЙ СЕТИ И ВЫЯВЛЕНИЯ КОМПЬЮТЕРНЫХ АТАК НА ПАКЕТНЫХ ВЫБОРКАХ ОПТИМАЛЬНОЙ ДЛИНЫ

А.А. Белоус, М.А. Маслова

В работе рассматривается актуальная задача повышения эффективности обнаружения киберугроз, реализуемых через сетевые каналы связи, с использованием метода оценки статистических характеристик сетевого трафика. Основное внимание уделено методу выбора оптимальной длины пакетной выборки, которая позволяет значительно сократить время, необходимое для выявления аномалий, путём минимизации временных затрат на обработку данных. В качестве ключевого критерия оптимальности предложено свойство корреляции признаков с целевой меткой: оптимальная выборка демонстрирует уровень корреляции, превосходящий корреляцию признаков для слишком коротких и слишком длинных выборок. Представленный алгоритм включает этапы по получению эталонных дампов трафика, выбору сетевых признаков, расчёту статистических характеристик и последовательному анализу корреляционных зависимостей с целью определения оптимальной выборки.

Введение

Основой эффективного противодействия инцидентам ИБ является непрерывный мониторинг сетевых событий [1]. В виду постоянно развивающихся угроз в области кибербезопасности, эффективное обнаружение аномального трафика в корпоративных сетях становится критически важным элементом обеспечения безопасности информационных ресурсов предприятий [2].

Согласно данным ГК InfoWatch [3], в 2023 году мировой показатель утечек конфиденциальной информации увеличился более чем на 60%. Общий объем скомпрометированных персональных данных вырос более чем в два раза, при этом крупные утечки (свыше 1 миллиона записей каждая) увеличились почти на 50%. Почти 95% всех зарегистрированных случаев утечек произошли вследствие кибератак, при этом практически все утечки связаны с сетевым каналом — их доля составляет около 98%.

В настоящее время можно выделить два укрупнённых метода, используемых для обнаружения компьютерных атак: сигнатурный и статистический. Данные методы анализа сетевого трафика обладают следующими недостатками:

1. Сигнатурные методы сохраняют зависимость от доступных на безвозмездной или коммерческой основе баз данных угроз и сигнатур, требующих регулярного обновления как со стороны поставщиков таких баз, так и со стороны пользователей; также данный вид анализа неэффективен против неизвестных или изменённых вариантов угроз.

2. Статистические методы, в отличие от сигнатурных, не требуют каких-либо баз данных — вместо этого в данных подходах оценивают локальные статистические параметры сети. Однако параметры подсчитываются на пакетной выборке, образуемой сетевой сессией, что в свою очередь требует верного поиска начала и конца сессии, а также ожидания накопления пакетов за весть период соединения. Заметим также, что сессии не имеют постоянной длины, что при каждом пересчёте статистик влечёт к изменению условий пересчёта, шумам и ошибкам в интерпретации данных.

Непредсказуемость длительности сессии даёт избыточное время злоумышленнику для реализации атаки и кражи чувствительной информации до того, как будет завершено соединение, определены границы сессии и пересчитаны сетевые признаки.

В настоящей работе предложен метод, предлагающий способ регулярного пересчёта сетевых характеристик без необходимости ожидания завершения сетевой сессии в режиме реального и близкого к реальному масштабу времени.

Анализ релевантных работ

Вопросы статистического анализа компьютерных сетей с целью выявления киберугроз находятся в центре внимания исследователей в последние годы. К настоящему моменту существует большое количество работ относящихся к данной тематике. Все они имеют место быть и использоваться для работ в последующих исследованиях.

В работе [4] берётся во внимание важность создания наборов данных, которые включают актуальные виды киберугроз учитывающие динамические изменения, касающиеся данного направления. Здесь важна область сетевых технологий, знания и навыки в области моделирования компьютерных атак и навыки практической работы с настройкой лабораторных стендов. Авторами работ были проанализированы и подобраны требования к наборам данных для обучения систем безопасности, такие как: конфигурация сети и её полнота, разновидности атак, виды используемых протоколов и обеспечение всестороннего сетевого взаимодействия. В результате исследования был получен набор данных CICIDS2017. Но данное исследование имеет минусы: не включает результаты тестирования методики в корпоративных сетях и не рассматривает возможные риски снижения эффективности обнаружения аномалий ввиду несоответствия характеристик сети, в которой происходил сбор данных, и защищаемой сети. Кроме того, сетевые характеристики в данном исследовании были подсчитаны на выборках, формируемых сетевыми сессиями переменной длительности.

Работа [5] является логическим продолжением работы [4], в котором разработана методика формирования обучающего набора для создания эффективной объектно ориентированной модели выявления компьютерных атак. Суть метода состоит в сборе данных непосредственно в реальной сети, применительно к защищаемому объекту или его аналогу, который полностью отражает все его особенности. Методика предусматривает использование различных программных инструментов для моделирования сетевого трафика с учётом временных интервалов и дополнительных параметров, что позволяет разнообразить реализацию вредоносных воздействий с учётом высокоточной имитации легитимных действий. Однако в работе большое внимание уделяется методу создания качественного набора данных, ввиду чего отсутствуют какие-либо предложения по оптимизации подсчёта статистических параметров на периоде, отличном от периода сетевой сессии.

В работе [6] предложен альтернативный подход к выявлению аномалий в сетевом трафике, который основан на оценке свойства самоподобия и применении статистических методов через подсчёт и сравнение с эталонным значением показателя Херста для параметров сессии. На первом шаге алгоритма рассматривается эталонный трафик, для которого рассчитывается показатель Херста. Далее уже анализируемый трафик делится на отрезки фиксированной временной длинны, для каждого из которых определяется свой коэффициент Херста. В случае значительного отклонения этого показателя от эталона делается вывод о наличии или отсутствии аномалий. В конце алгоритма производится анализа показателей для точного выявления векторов атаки. В завершение исследования проведён анализ статистических и фрактальных методов, после чего выбраны наиболее эффективные алгоритмы для реализации предложенного подхода. В качестве расчётного интервала в работе используются временные отрезки по 120 секунд, однако оптимальность выбора данных отрезков не доказана через сравнение с альтернативными временными интервалами, которые могли бы быть использованы в предложенном авторами методе, что также не исключает угрозу избыточности времени определения аномалий, чем может воспользоваться злоумышленник.

Таким образом, несмотря на значительные достижения в разработке методик выявления киберугроз путём оценки статистических характеристик сети, проблема поиска универсального расчётного интервала для анализа сетевого трафика остаётся нерешенной.

Описание исследования

Для проведения исследования был выбран набор данных CIC-IDS2017 [7] ввиду хорошего разнообразия представленных угроз, а также большей актуальности данных в сравнении с аналогичными открытыми наборами. В условиях базового исследования был выбран десяток сетевых характеристик, которые показали наилучшую корреляцию с целевой меткой в смежных работах [8, 9]. Перечень выбранных характеристик, а также комментарии к ним, представлен в табл. 1

Таблица 1

Применяемые при расчётах статистические характеристики сети

Наименование характеристики	Пояснение	Сокращение, используемое в настоящей работе
Average Packet Size	коэффициент средней длины полезной нагрузки кадров	APS
Flow Bytes/s	показатель скорости движения кадров	FB
Max Packet Length	показатель максимальной длина кадра	MPL
Fwd Packet Length Mean	коэффициент средней длина кадров, переданных в направлении к целевому хосту	FPLMean
Fwd IAT Min	показатель минимального значения интервала между кадрами, переданными к целевому хосту	FwdIM
Total Length of Fwd Packets	показатель общей длины кадров, переданных к целевому хосту	TLFP
Fwd IAT Std	коэффициент стандартного отклонения значения интервала между кадрами	FIS
Flow IAT Mean	коэффициент среднего значения интервала между кадрами	FlowIM
Fwd Packet Length Max	показатель максимальной длины кадра, переданного к целевому хосту	FPLMax
Fwd Header Length	величина общей длины заголовков кадров, переданных к целевому хосту	FHL

Перечень атак, отобранных для расчёта статистических характеристик и подготовки данных для исследования действенности предлагаемого метода, представлен в табл. 2.

Таблица 2

Описание исследуемых атак из CIC-IDS2017

Тип атаки	Сокращение, используемое в настоящей работе	Продолжительность атаки, чч:мм:сс	Число легитимных пакетов, ед.	Число аномальных пакетов, ед.
Botnet ARES	BA	01:00:00	121 255	136 260
Brute Force over HTTP	BF	00:40:00	10 856	39 145
Cross-Site Scripting	XSS	00:20:00	5 724	14 943

Из табл. 2 можно сделать вывод о таком недостатке набора CIC-IDS2017, как плохая балансировка данных по классам атак и числу легитимных и вредоносных пакетов: атака BA длилась 60 минут, в течение которых накопилось более ста тысяч пакетов каждого типа, в то время как атака XSS длилась 20 минут, за которые в сумме накопилось чуть

более двадцати тысяч пакетов — суммарное число пакетов для рассматриваемых атак отличается на порядок, что может затруднить подсчёт статистики на пакетных выборках не оптимально большого размера. Далее учтём проблему некачественной балансировки данных при оценке их статистических свойств.

Для каждой атаки выбранные ранее статистические характеристики были пересчитаны на пакетных выборках длиной в 10, 50, 100, 250, 500, 750, 1 000, 5 000 и 10 000 пакетов, результаты подсчётов сохранены в csv файлы для нормального и атакующего трафика отдельно.

Далее для каждой пары файлов, содержащих характеристики, подсчитанные на одном и том же пакетном интервале, построены диаграммы рассеяния полученных значений нормального (круглые чёрные o-точки) и атакующего (крестообразные серые x-точки) трафика, собранного между фиксированными сетевыми узлами, в зависимости от длины пакетной выборки. На оси X отложены значения сетевого параметра, на оси Y — индексы, которые каждое новое значение получает при записи в csv-файл; данное группирование значений вдоль оси Y позволяет отразить кучность точек без существенного взаимного перекрытия и искажении информации о взаимном распределении.

Также пунктирной линией светло-серого цвета проведена граница, представляющая собой геометрическое место точек, которые являются средним арифметическим между медианами распределений чёрных и серых точек, что позволяет наглядно наблюдать разделение между двумя классами.

Далее были подсчитаны отношения числа атакующих точек к числу нормальных точек слева и справа от границы, по которым можно осуществить грубую количественную оценку разделимости признаков по классам легитимного и вредоносного трафика.

С учётом вышеупомянутой проблемы несбалансированности данных по классам атак и типам пакетов, дополнительно проведена оценка корреляции между признаками и целевой меткой (наличие или отсутствие аномалии) с использованием методов оценивания, более устойчивых к шумам, чем граница, проходящая через медианное среднее распределений нормальных и вредоносных пакетов. В базовом исследовании принято допущение о существовании линейной зависимости между признаками и целевой меткой, на основании чего выбран коэффициент корреляции Пирсона.

Коэффициент корреляции Пирсона k_{XY} для двух множеств X, Y может быть рассчитан по формуле:

$$k_{XY} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

где n – размер выборки;

x_i, y_i – i-е точки множеств X, Y.

Коэффициент корреляции Пирсона принимает значение в пределах от -1 до 1, где -1 означает идеальную обратную корреляцию, а 1 — идеальную прямую зависимость. В данной работе информативными признаками считаются те, которые демонстрируют значение корреляции по модулю не менее 0,7.

На рис. 1 и 2 представлены примеры диаграмм, построенных для такого признака, подсчитанного на выборках в 10 и 750 TCP/IP пакетов, как средняя длина пакетов, переданных в прямом направлении, тип атаки: Botnet ARES.

На рис. 1 видно, что нельзя выделить какой-либо интервал слева и справа от границы разделения таким образом, чтобы в пределах данного интервала преобладали точки одного типа, а вне данного интервала — точки другого типа. Это подтверждают подсчёты отношений атакующих к нормальным пакетам по обе стороны от границы разделения — 1,12:1,13 что приблизительно можно считать за 1:1. Выборки длиной в 10 пакетов недостаточно для наглядного разделения трафика.

В случае рис. 2, по подсчитанным на выборках в 750 пакетов значениям рассматриваемой характеристики явно видно, что слева от границы преобладают o-точки, принадлежащие нормальному трафику.

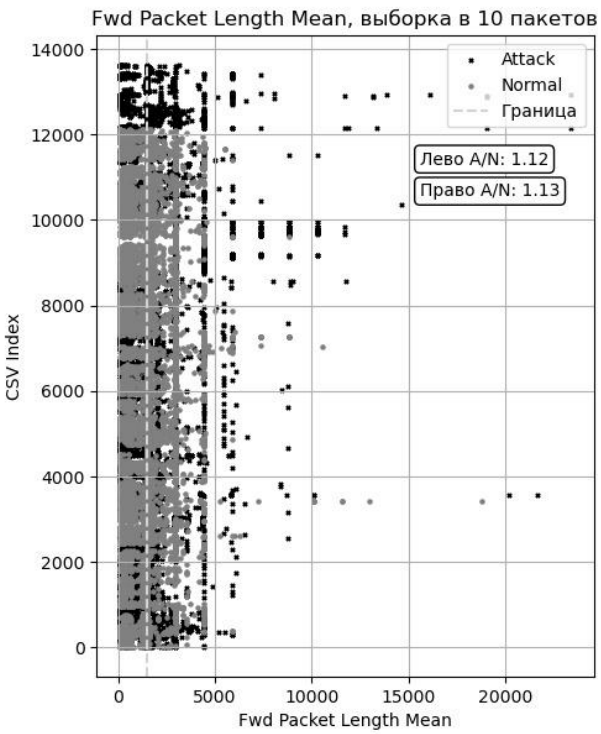

Рис. 1. Распределение FPLMean, подсчитанных на выборках в 10 пакетов, атака BA

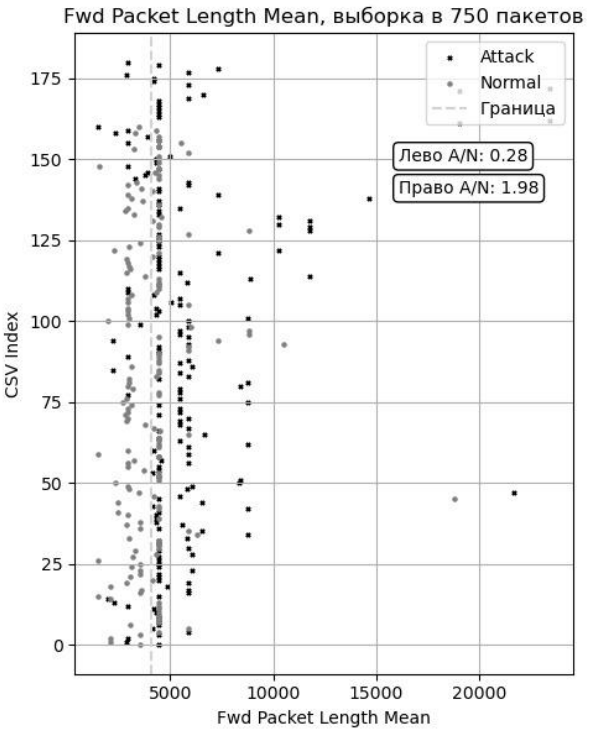

Рис. 2. Распределение FPLMean, подсчитанных на выборках в 750 пакетов, атака BA

Также приведённые на рис. 2 расчёты показывают преобладание нормальных точек слева от границы более чем в 7 раз в сравнении с правой полуплоскостью. Из рис. 1 и 2 можно сделать вывод, что выборка в 750 пакетов является оптимальной, однако для подтверждения оптимальности следует провести более помехоустойчивые расчёты.

Коэффициенты корреляции Пирсона между всеми рассматриваемыми статистическими признаками и целевой меткой для атаки BA представлены в табл. 3. Проверим, подтвердят ли данные в табл. 3 предположение об оптимальности выборки в 750 пакетов.

Таблица 3

Коэффициенты корреляции Пирсона между признаками и целевой меткой для атаки BA

Длина пакетной выборки, число пакетов	Статистические характеристики сети, см. табл. 1									
	APS	FB	FlowIM	FHL	FwdIM	FIS	FPLMax	FPLMean	MPL	TLFP
10	0.03	0.04	0.015	0.058	0.017	0.016	0.057	0.057	0.03	0.03
50	0.034	0.058	0.033	0.07	0.04	0.024	0.101	0.101	0.034	0.034
100	0.036	0.051	0.039	0.076	0.026	0.03	0.127	0.127	0.036	0.036
250	0.037	0.052	0.037	0.084	0.019	0.032	0.195	0.195	0.037	0.037
500	0.039	0.062	0.049	0.103	0.001	0.046	0.261	0.261	0.039	0.039
750	0.041	0.064	0.033	0.113	0.115	0.052	0.301	0.301	0.04	0.04
1 000	0.042	0.067	0.048	0.121	0.153	0.052	0.304	0.304	0.041	0.041
5 000	0.052	0.181	0.017	0.265	0.259	0.087	0.437	0.437	0.052	0.052
10 000	0.05	0.194	0.117	0.366	0.243	0.168	0.461	0.461	0.05	0.05

Из табл. 3 видно, что для параметра FPLMean на самом деле отсутствует наглядная зависимость с целевой меткой вплоть до выборки в 10 000 пакетов. Дальнейшее увеличение пакетной выборки может быть не целесообразно ввиду возникновения проблем ожидания накопления пакетов и усложнения вычислений, на решение которых направлено настоящее исследование.

На рис. 3 и 4 показано облако распределения стандартного отклонения межпакетного интервала для нормального и атакующего трафика, подсчитанного на выборках в 100 и 250 TCP/IP пакетов соответственно, тип атаки: Brute Force.

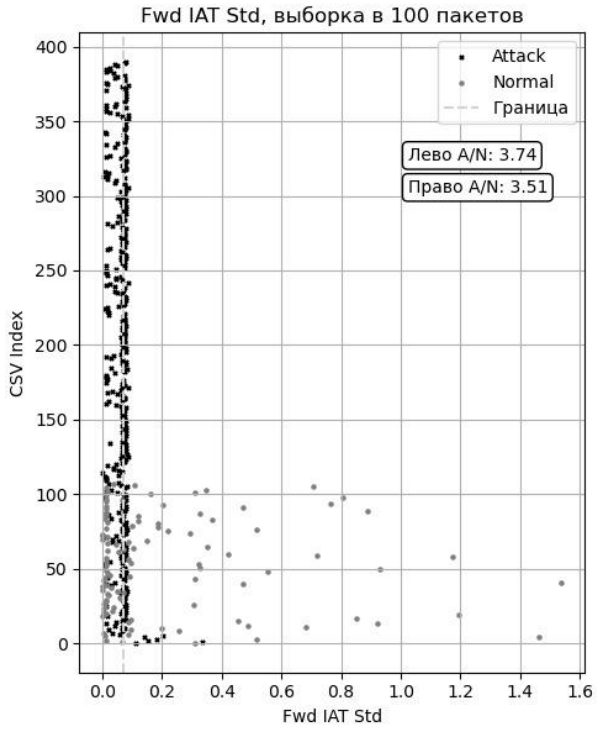

Рис. 3. Распределение FIS, подсчитанных на выборках в 100 пакетов, атака BF

На рис. 3 пропорции между атакующим и нормальным трафиком слева и справа от границы приблизительно равны, исходя из чего можно сделать вывод о недостаточно наглядной разделимости трафика при оценке данной характеристики, подсчитанной на выборках в 100 пакетов, применимо к атаке рассматриваемого типа.

На рис. 4, напротив, заметно сильное преобладание x-точек слева от границы разделения, что также подтверждают расчёты: слева от границы аномальных значений рассматриваемого признака более чем в 157 раз больше, чем справа. Также можно видеть, что x-точки слева от границы практически не превосходят некоторого постоянного значения, примерно равного 0,08.

Можно сделать промежуточный вывод о разделимости трафика на аномальный и легитимный при оценке параметра FIS, подсчитанного на выборках в 250 пакетов, для атаки BF.

Коэффициенты корреляции Пирсона пересчитанный заново между всеми рассматриваемыми статистическими признаками и целевой меткой для атаки Brute Force представлены в табл. 4. Проведём анализ полученных данных.

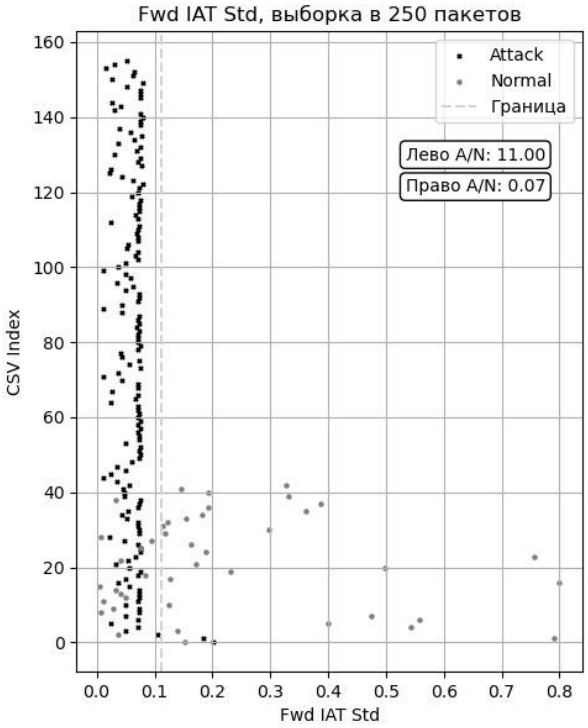

Рис. 4. Распределение FIS, подсчитанных на выборках в 250 пакетов, атака BF

Из табл. 4 видно, что, начиная с выборки в 500 пакетов, некоторые статистические признаки коррелируют с целевой меткой на уровне 0,7 и более, а на выборках в 5 000 и 10 000 пакетов уже видно 7 признаков, которые коррелируют на уровне 0,9 и более. Также на выборке в 10 000 пакетов уровень корреляции сравним с уровнем для 5 000 пакетов, из чего следует, что нет необходимости накапливать 10 000 пакетов — выборка в 5 000 накапливается быстрее,

вследствие чего требует меньше ресурсов для расчётов и является оптимальной.

Таблица 4

Коэффициенты корреляции Пирсона между признаками и целевой меткой для атаки BF

Длина пакетной выборки, число пакетов	Статистические характеристики сети, см. табл. 1									
	APS	FB	FlowIM	FHL	FwdIM	FIS	FPLMax	FPLMean	MPL	TLFP
10	0.176	0.161	0.122	0.316	0.35	0.113	0.268	0.272	0.176	0.176
50	0.353	0.166	0.316	0.357	0.339	0.273	0.187	0.182	0.353	0.353
100	0.473	0.139	0.464	0.388	0.224	0.398	0.336	0.33	0.473	0.473
250	0.576	0.005	0.641	0.476	0.295	0.548	0.385	0.376	0.575	0.575
500	0.698	0.396	0.751	0.598	0.299	0.653	0.329	0.316	0.697	0.697
750	0.754	0.773	0.845	0.69	0.24	0.784	0.312	0.297	0.753	0.753
1 000	0.776	0.654	0.825	0.684	0.213	0.714	0.286	0.263	0.774	0.774
5 000	0.964	0.98	0.987	0.964	-	0.995	0.156	0.188	0.964	0.964
10 000	0.993	0.989	0.998	0.996	-	0.998	0.333	0.333	0.993	0.993

На рис. 5 и 6 показано облако распределения минимального межпакетного интервала для нормального и атакующего трафика, подсчитанного на выборках в 100 и 10 TCP/IP пакетов соответственно, тип атаки: XSS.

В случае рис. 5, подсчитанных на интервалах в 100 пакетов точек недостаточно для грамотного построения границы — пропорции атакующих и нормальных значений слева и справа от медианного среднего примерно равны.

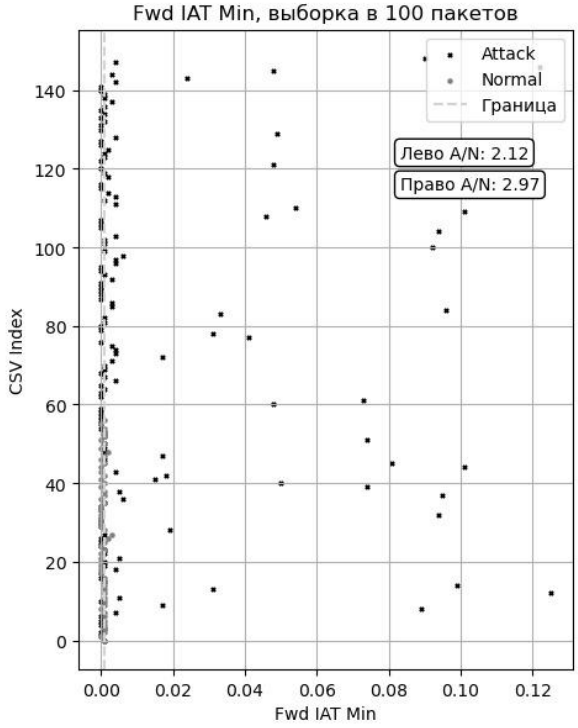

Рис. 5. Распределение Fwd IAT Min, подсчитанных на выборках в 100 пакетов, XSS

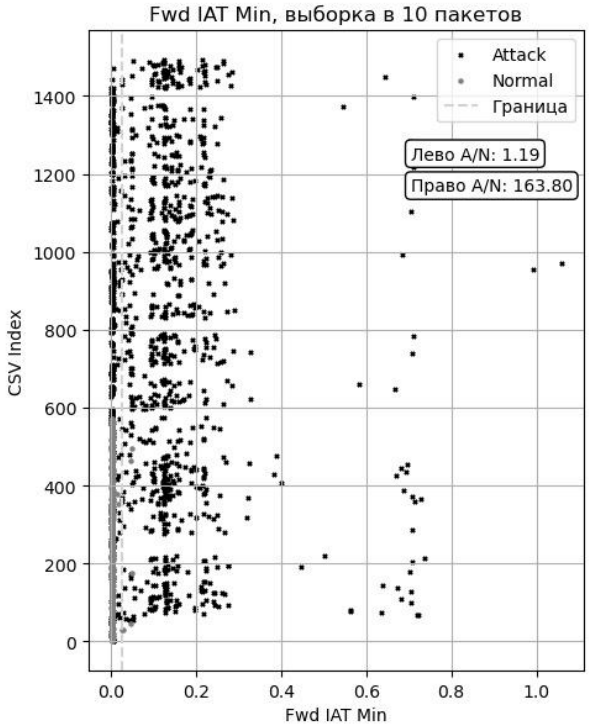

Рис. 6. Распределение Fwd IAT Min, подсчитанных на выборках в 10 пакетов, XSS

На рис. 6 граница разделяет трафик более наглядно, а пропорции между атакующими и нормальными точками до и после границы отличаются более чем в 164 раза. Выборка в 10 пакетов для данной характеристики является оптимальной.

Коэффициенты корреляции Пирсона между рассматриваемыми статистическими признаками и целевой меткой для данного типа атаки представлены в табл. 5. Оценим, подтверждают ли данные в табл. 5 полученные ранее результаты.

Таблица 5

Коэффициенты корреляции Пирсона между признаками и целевой меткой для атаки XSS

Длина пакетной выборки, число пакетов	Статистические характеристики сети, см. табл. 1									
	APS	FB	FlowIM	FHL	FwdIM	FIS	FPLMax	FPLMean	MPL	TLFP
10	0.358	0.061	0.224	0.256	0.351	0.194	0.477	0.479	0.358	0.358
50	0.531	0.046	0.4	0.282	0.331	0.276	0.49	0.496	0.53	0.53
100	0.588	0.155	0.55	0.301	0.25	0.33	0.389	0.399	0.588	0.588
250	0.647	0.348	0.687	0.377	0.148	0.442	0.127	0.14	0.647	0.647
500	0.752	0.815	0.829	0.496	0.14	0.628	0.001	0.013	0.752	0.752
750	0.785	0.826	0.839	0.614	0.175	0.624	0.001	0.02	0.785	0.785
1 000	0.867	0.856	0.88	0.709	0.141	0.704	0.112	0.123	0.867	0.867
5 000	0.993	0.985	0.988	1	-	0.993	0.475	0.5	0.993	0.993

Из табл. 5 видно, что представленного числа пакетов (чуть более двадцати тысяч) крайне мало для формирования достаточно больших множеств нормальных и аномальных значений, если подсчитывать их на выборках в 10 000 пакетов; также, начиная с 500 пакетов, в каждой записи можно выделить пять и более признаков, коррелирующих с целевой меткой на уровне 0,7 и более, а на выборках в 1 000 и 5 000 пакетов корреляция некоторых признаков улучшается с 0,8 до 0, 9. Таким образом. выборка в 5 000 пакетов представляется оптимальной для выявления данного типа атаки в контексте рассматриваемых сетевых характеристик.

Из представленных на рис. 1-6, а также в табл. 3-5 данных можно сделать промежуточный вывод о нелинейности связи между статистическими параметрами сети и ответом на вопрос о наличии или отсутствии аномалий в каждый момент времени; также из материалов видно, что для различных сетевых атак длина пакетной выборки, оптимальной для наглядного разделения трафика, отличается.

Заключение

В свете постоянно растущих и развивающихся угроз безопасности объектов информатизации актуальным вектором атак злоумышленников остаются угрозы, реализуемые через сетевой канал связи.

Одним из наиболее перспективных методов оперативного обнаружения киберугроз является метод оценки статистических характеристик сети и их совокупного сравнения с эталонным значением. Несмотря на достижения в этой области, существует проблема поиска универсального расчётного интервала, который минимизировал бы время, необходимое для подготовки такого интервала к анализу, тем самым уменьшая время активных действий злоумышленника до того момента, когда система защиты накопит достаточно информации, чтобы сделать предположение о наличии вредоносной активности.

В аналогичных исследованиях пакетная выборка формируется либо на основе интервала, обозначенного сетевой сессией непредсказуемой длительности, либо на основе интервала, обозначенного фиксированным временным отрезком, оптимальная продолжительность которого не аргументирована в явном виде.

В настоящей работе представлен метод оценки сетевых характеристик, в котором в качестве расчётного интервала берётся

пакетная выборка оптимальной длинны. Под оптимальностью в исследовании понимается свойство подсчитанных статистических характеристик, которое заключается в наличии уровня корреляции с целевой меткой, сравнимого с корреляцией признаков, подсчитанных на выборках чуть меньшей или чуть большей длины, но явно превосходящего уровень корреляции признаков, подсчитанных на выборках существенно меньшей или существенно большей длины.

В процессе исследования по результатам смежных работ выбраны статистические параметры, а также три типа атак из открытого набора CICIDS2017, принадлежащие к различным классам киберугроз. Далее подсчитаны эталонные и вредоносные статистические показатели для каждой из рассматриваемых атак, после чего произведена попытка линейного разделения данных с целью решения задачи бинарной классификации трафика на аномальный и легитимный; результаты поиска оптимальной длины пакетной выборки проверены с помощью менее чувствительного к выбросам и плохой балансировке по классам метода Пирсона. В результате можно сделать вывод об адекватности и перспективности представленного метода.

Формализовать алгоритм предложенного метода можно следующим образом:

1. Произвести в защищаемой сети самостоятельное тестирование на проникновение с целью сбора вредоносного трафика, учитывающего особенности сетевых ресурсов.

2. Получить эталонные дампы нормального и скомпрометированного трафика.

3. Выбрать экспериментальные сетевые признаки для подсчёта.

4. Выбрать экспериментальные значения длины пакетных выборок, на которых будут производиться расчёты.

5. Рассчитать эталонные и аномальные статистические признаки на всех пакетных выборках.

6. Задать пороговое значение для уровня корреляции признака с целевой меткой.

7. Для атаки каждого типа, актуального для защищаемого ресурса, последовательно оценивать совокупные значения коэффициентов корреляции в сравнении с пороговым значением для каждой выборки; первая выборка наименьшей длинны, удовлетворяющая вышеуказанному принципу оптимальности — искомая.

Дальнейшее направление исследования может быть связано с созданием набора данных, в котором будут частично или полностью убраны недостатки используемого в работе набора; рассмотрением большего числа атак различных классов киберугроз; увеличением пространства сетевых статистических характеристик, используемых в исследовании; применением передовых технологий машинного обучения с целю выявления нелинейных зависимостей между признаками и целевой переменной, а также с целью решения задачи уже многоклассовой классификации, т.е. поиска пакетной выборки такой длины, анализ подсчитанных на которой статистических параметров позволит определить тип атаки среди возможных классов.

Список литературы

1. Рытов М.Ю. Порядок оценки уровня эффективности системы непрерывного противодействия инцидентам информационной безопасности на объекте / М.Ю. Рытов, О.М. Голембиовская, Е.В. Кондрашова [и др.] // Информация и безопасность. 2024. Т. 27. Вып. 1. С. 135-142.

2. Maslova, M. Analysis and assessment of information security risks for sustainable development of the oil and gas industry / M. Maslova // E3S Web of Conferences : II International Conference on Environmental Technologies and Engineering for Sustainable Development (ETESD-II 2023), Tashkent, 13–15 сентября 2023 года. Vol. 443. Tashkent: EDP Sciences, 2023. P. 06010.

3. Экспертно-аналитический центр ГК InfoWatch. Аналитический отчёт. Утечки информации в мире, 2022-2023 годы. URL: https://www.infowatch.ru/sites/default/files/analytics/files/issledovaniye-utechek-informatsii-v-mire-za-2022-2023-gody.pdf (дата обращения: 28.10.2024).

4. Sharafaldin I., Lashkari A.H., Ghorbani Ali A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic

Characterization. In Proc. of the 4th International Conference on Information Systems Security and Privacy (ICISSP). 2018. P. 108-116.

5. Гетьман А.И. Методика сбора обучающего набора данных для модели обнаружения компьютерных атак. / А.И. Гетьман, М.Н. Горюнов, А.Г. Мацкевич, Рыболовлев Д.А. // Труды ИСП РАН. Т 33. Вып. 5. 2021. С. 83-104.

6. Веселова, В.А. Подход к обнаружению аномалий в самоподобном сетевом трафике / В.А. Веселова, В.С. Коломойцев // Надежность. 2023. Т. 23. № 2. С. 57-63.

7. Intrusion Detection Evaluation Dataset (CIC-IDS2017). URL: https://www.unb.ca/cic/datasets/ids-2017.html (дата обращения: 28.10.2024).

8. Leevy, Joffrey & Hancock, John & Zuech, Richard & Khoshgoftaar, Taghi. (2021). Detecting cybersecurity attacks across different network features and learners. Journal of Big Data. URL: https://www.researchgate.net/publication/349546005_Detecting_cybersecurity_attacks_across_different_network_features_and_learners (дата обращения: 28.10.2024).

9. Горюнов, М. Н. Синтез модели машинного обучения для обнаружения компьютерных атак на основе набора данных CICIDS2017 / М. Н. Горюнов, А. Г. Мацкевич, Д. А. Рыболовлев // Труды Института системного программирования РАН. 2020. Т. 32, № 5. С. 81-94.

Севастопольский государственный университет
Sevastopol State University

Поступила в редакцию 5.11.24

Информация об авторах

Маслова Мария Александровна – доцент кафедры “Информационная безопасность”, Севастопольский государственный университет, e-mail: mashechka-81@mail.ru

Белоус Александр Александрович – студент четвёртого курса кафедры «Информационная безопасность» Института информационных технологий, Севастопольский государственный университет, e-mail: belous.a.a.m@yandex.ru

A METHOD FOR STATISTICAL ANALYSIS OF A LOCAL NETWORK AND DETECTION OF COMPUTER ATTACKS ON PACKET SAMPLES OF OPTIMAL LENGTH

A.A. Belous, M.A. Maslova

The paper considers the urgent task of increasing the effectiveness of detecting cyber threats implemented through network communication channels using the method of evaluating the statistical characteristics of network traffic. The main attention is paid to the method of choosing the optimal batch sampling length, which significantly reduces the time required to identify anomalies by minimizing the time spent on data processing. The property of correlation of features with the target label is proposed as a key criterion of optimality: the optimal sample demonstrates a level of correlation that exceeds the correlation of features for too short and too long samples. The presented algorithm includes steps for obtaining reference traffic dumps, selecting network features, calculating statistical characteristics and sequentially analyzing correlation dependencies in order to determine the optimal sample.

Submitted 5.11.24

Information about the author

Maria A. Maslova – Associate Professor of the Department of Information Security, Sevastopol State University, e-mail: mashechka-81@mail.ru

Aleksandr A. Belous – fourth-year student of the Information Security Department of the Institute of Information Technologies, Sevastopol State University, e-mail: belous.a.a.m@yandex.ru