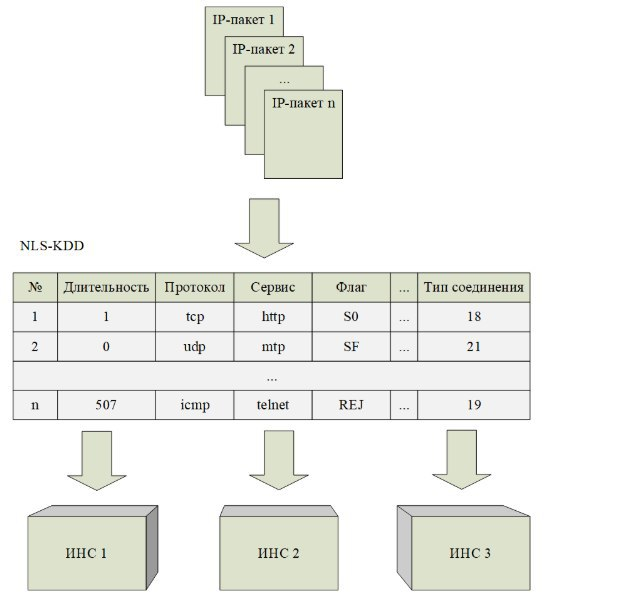
**Чекпоинт 2**

1. Архитектуру системы с агентами, сервером, отбором признаков классификации на агентах

Архитектура клиент (Агент + служба мониторинга и индикации) – сервер (нейросети, анализ и мониторинг).

* Агент установлен на клиенте, связь и целостность агента контролируется сервером по закрытому протоколу
* Агент собирает данные и отправляет на сервер



В таком виде с Агентов трафик поступает на сервер, то есть Агенты проводят предварительную подготовку основных признаков для анализа

* На сервере производится анализ, обученная нейросеть выдает комплексное заключение через 200 мс
* Данные анализа нейросети возвращаются Агенту, тот в свою очередь подтверждает получение и производит индикацию
* Индикация сообщений производится с помощью отдельной службы, которая мониторит деятельность Агента и сообщений сервера, так же по закрытому протоколу
* В случае не обнаружения атаки, индикатор состояния остается зеленым
* В случае подозрения на атаку, вероятность выше 55%, индикатор меняет цвет на оранжевый, выводится сообщение на дисплее Администратора сети (подтверждение прочтения не требуется)
* В случае обнаружения атаки, вероятность выше 65%, индикатор меняет цвет на красный, сообщение получает Администратор сети и пользователь. В индикации рекомендации для пользователя – сохранить все данные и закрыть все программы, не производить никаких действий до изменения индикатора в зеленый цвет или до сообщения администратора.
* В случае неполучения подтверждения от Агента, а также службы мониторинга и индикации), сервер отправляет сообщение Администратору сети.
* Для критических случаев - если администратор сети не подтвердил получение сообщения, сервер совершает звонок на мобильные телефоны и отправляет сообщение в мессенджеры и на электронную почту.

Параллельно с основной работающей нейросетью работает вторая нейросеть и обучается, отображая новые признаки и классы поведения сети.?

**2. Базы данных обучения с описанием**

Обычно для обучения и тестирования ИНС имеющиеся экспериментальные данные разбиваются на обучающую БД и контрольную БД. В проведённых исследованиях в качестве обучающей БД с параметрами сетевых соединений, представляющая собой 10 % от исходного **CSE-CIC-IDS2018**. Записи из полной контрольной БД подавались на обученные НС. На основании результатов этого этапа определяется статистика и даётся анализ эффективности НС по критериям качества распознавания типов аномальных соединений и наличию ложных срабатываний (когда нормальное соединение принимается за атаку).

|  |  |
| --- | --- |
| Название | kddcup99 |
| Ссылка | http://kdd.ics.uci.edu/databases/kddcup99/ |
| Описание | Данная БД преследовала цель построения определителя сетевых атак. Модель различала «хорошие», «нормальные» и «плохие» (вторжение) сетевые соединения. Данная БД содержит множество вторжений, эмулированных с помощью военной сети. |
| Кол-во атак | БД содержит 24 различных типа атак, с дополнительными 14 видами исключительно в тестовых данных |
| Типы атак | back dos  buffer\_overflow u2r  ftp\_write r2l  guess\_passwd r2l  imap r2l  ipsweep probe  land dos  loadmodule u2r  multihop r2l  neptune dos  nmap probe  perl u2r  phf r2l  pod dos  portsweep probe  rootkit u2r  satan probe  smurf dos  spy r2l  teardrop dos  warezclient r2l  warezmaster r2l  Attacks fall into four main categories:  DOS: denial-of-service, e.g. syn flood;  R2L: unauthorized access from a remote machine, e.g. guessing password;  U2R: unauthorized access to local superuser (root) privileges, e.g., various ``buffer overflow'' attacks;  probing: surveillance and other probing, e.g., port scanning. |

|  |  |
| --- | --- |
| Название | SNMP-MIB Dataset |
| Ссылка | https://www.researchgate.net/publication/322887764\_Towards\_Generating\_Realistic\_SNMP-MIB\_Dataset\_for\_Network\_Anomaly\_Detection |
| Описание | Данные были собраны на основании Management Information Base (MIB) |
| Кол-во атак | БД включает в себя 4998 записей, каждая из которых состоит из 34 MIB переменных, которые в последующем категорируются с соответствующие группы по именам: Interface, IP, TCP and ICMP. |
| Типы атак | DOS: denial-of-service, e.g. syn flood;  R2L: unauthorized access from a remote machine, e.g. guessing password;  U2R: unauthorized access to local superuser (root) privileges, e.g., various ``buffer overflow'' attacks; |

|  |  |
| --- | --- |
| Название | CIRA-CIC-DoHBrw-2020 |
| Ссылка | <https://www.unb.ca/cic/datasets/dohbrw-2020.html> |
| Описание | Двухуровневый подход используется для захвата доброкачественного и вредоносного DoH-трафика вместе с non-DoH трафиком. |
| Кол-во атак |  |
| Типы атак | Для создания репрезентативного набора данных HTTPS и DoH-трафик генерируется путем доступа к 10 тысячам лучших веб-сайтов Alexa и с использованием браузеров и инструментов туннелирования DNS, которые поддерживают протокол DoH соответственно. На первом уровне захваченный трафик классифицируется как DoH и non-DoH с помощью классификатора статистических функций. На втором уровне трафик DoH характеризуется как доброкачественный DoH и вредоносный DoH с помощью классификатора временных рядов. |

|  |  |
| --- | --- |
| Название | CICDDoS2019 |
| Ссылка | <https://www.unb.ca/cic/datasets/ddos-2019.html> |
| Описание | CICDDoS2019 содержит безопасные и самые современные распространенные DDoS-атаки, которые напоминают настоящие данные реального мира (PCAP). Он также включает результаты анализа сетевого трафика с использованием CICFlowMeter-V3 с помеченными потоками на основе метки времени, IP-адресов источника и назначения, портов источника и назначения, протоколов и атак (файлы CSV). |
| Кол-во атак | 19 DDoS-атак |
| Типы атак | Для этого набора данных было взято абстрактное поведение 25 пользователей на основе протоколов HTTP, HTTPS, FTP, SSH и электронной почты. Было выполнено 12 DDoS-атак, включая NTP, DNS, LDAP, MSSQL, NetBIOS, SNMP, SSDP, UDP, UDP-Lag, WebDDoS, SYN и TFTP, в день обучения и 7 атак, включая PortScan, NetBIOS, LDAP, MSSQL, UDP, UDP. -Лаг и SYN в день тестирования. Объем трафика для WebDDoS был настолько низким, что PortScan только что был запущен в день тестирования и будет неизвестен для оценки предложенной модели. |

|  |  |
| --- | --- |
| Название | VPN-nonVPN dataset (ISCXVPN2016) |
| Ссылка | <https://www.unb.ca/cic/datasets/vpn.html> |
| Описание | Чтобы создать репрезентативный набор данных реального трафика в ISCX, был определён набор задач, гарантируя, что условный набор данных достаточно богат разнообразием и количеством. Были созданы учетные записи для пользователей Алисы и Боба, чтобы использовать такие сервисы, как Skype, Facebook и т. д. |
| Кол-во атак | Трафик был записан с помощью Wireshark и tcpdump, в результате чего было создано 28 ГБ данных. Для VPN был использован внешний поставщик услуг VPN и подключение к нему производилось с помощью OpenVPN (режим UDP). Для генерации трафика SFTP и FTPS был использован внешний поставщик услуг и Filezilla в качестве клиента. |
| Типы атак | 14 категорий трафика: VOIP, VPN-VOIP, P2P, VPN-P2P и т. д. |

|  |  |
| --- | --- |
| Название | URL dataset (ISCX-URL2016) |
| Ссылка | <https://www.unb.ca/cic/datasets/url-2016.html> |
| Описание | Был использован упрощенный подход к обнаружению и категоризации вредоносных URL-адресов в соответствии с их типом атаки и показано, что лексический анализ эффективен и действенен для упреждающего обнаружения этих URL-адресов. Было изучено влияние методов обфускации на вредоносные URL-адреса, чтобы определить тип метода обфускации, нацеленный на конкретный тип вредоносного URL. |
| Кол-во атак | С лучших веб-сайтов Alexa было собрано более 35 300 безопасных URL-адресов. Домены прошли через поисковый робот Heritrix для извлечения URL-адресов. Около полумиллиона уникальных URL-адресов сначала сканируются, а затем передаются для удаления повторяющихся URL-адресов и URL-адресов только для домена. Позже извлеченные URL-адреса были проверены через Virustotal, чтобы отфильтровать безопасные URL-адреса. |
| Типы атак | URL-адреса спама: около 12 000 URL-адресов спама было собрано из общедоступного набора данных WEBSPAM-UK2007.  Фишинговые URL-адреса: около 10 000 фишинговых URL-адресов были взяты из OpenPhish, который является хранилищем активных фишинговых сайтов.  URL-адреса вредоносных программ: более 11500 URL-адресов, связанных с вредоносными веб-сайтами, были получены из DNS-BH, проекта, который поддерживает список вредоносных сайтов.  URL-адреса Defacement: Более 45 450 URL-адресов относятся к категории URL-адресов Defacement. Это доверенные веб-сайты, получившие рейтинг Alexa, на которых размещены мошеннические или скрытые URL-адреса, содержащие обе вредоносные веб-страницы.  Back, buffer\_overflow, ftp\_write, guess\_passwd, Imap, ipsweep, Land, loadmodule, multihop, Neptune, Perl, Phf, Pod, postweep |

|  |  |
| --- | --- |
| Название | NSL-KDD dataset |
| Ссылка | <https://www.unb.ca/cic/datasets/nsl.html> |
| Описание | NSL-KDD - это набор данных, предложенный для решения некоторых проблем, присущих набору данных KDD'99. Количество записей в наборе NSL-KDD и тестовых наборах является разумным. Это преимущество делает возможным проведение экспериментов на полном наборе без необходимости случайного выбора небольшой части. Следовательно, результаты оценки различных исследовательских работ будут последовательными и сопоставимыми. |
| Кол-во атак | Оригинальные записи | Отчетливые записи | Скорость снижения  Attacks: 3,925,650 | 262,178 | 93.32%  Normal: 972,781 | 812,814 | 16.44%  Total: 4,898,431 | 1,074,992 | 78.05% |
| Типы атак | back dos  buffer\_overflow u2r  ftp\_write r2l  guess\_passwd r2l  imap r2l  ipsweep probe  land dos  loadmodule u2r  multihop r2l  neptune dos  nmap probe  perl u2r  phf r2l  pod dos  portsweep probe  rootkit u2r  satan probe  smurf dos  spy r2l  teardrop dos  warezclient r2l  warezmaster r2l  Attacks fall into four main categories:  DOS: denial-of-service, e.g. syn flood;  R2L: unauthorized access from a remote machine, e.g. guessing password;  U2R: unauthorized access to local superuser (root) privileges, e.g., various ``buffer overflow'' attacks;  probing: surveillance and other probing, e.g., port scanning. |

|  |  |
| --- | --- |
| Название | ADFA-WD (Windows dataset) |
| Ссылка | <https://dibbs.ai.arizona.edu/dibbs/adfa-ids/readme.txt> |
| Описание | ADFA-WD (набор данных Windows) был создан на хост-ОС Windows XP Service Pack 2 с включенным брандмауэром XP по умолчанию для всех атак, включенным общим доступом к файлам, настроенным сетевым принтером, беспроводной сетью и сетью Ethernet. Norton AV 2013 использовался для сканирования определенных полезных данных. Были активированы FTP-сервер, веб-сервер и инструмент управления, а также пакет цифрового радио потокового аудио. |
| Кол-во атак | Целевое соотношение 1: 10: 1  нормальные данные: данные проверки  данные атаки использовалось для управления сбором и структурированием действий. |
| Типы атак | Imap, ipsweep, Land, loadmodule, multihop, Neptune, Perl, Phf, Pod, postweep |

|  |  |
| --- | --- |
| Название | Aktaion Example Labeled Data |
| Ссылка | https://dibbs.ai.arizona.edu/dibbs/aktaion-example-labeled-data/Aktaion-ReadMe.txt |
| Описание | Эта БД содержит помеченные данные сетевого трафика в формате ARFF. Первоначальной целью было обучение обнаружению программ-вымогателей в Aktaion IDS. Данные относятся к августу 2016 года.  Этот набор данных представляет собой набор помеченных примеров для инструмента обнаружения программ-вымогателей на основе машинного обучения Aktaion. Aktaion можно классифицировать как IDS, основанную на аномалиях, и данные в этой коллекции используются в качестве примера для этого инструмента. |
| Кол-во атак | Трафик был записан с помощью Wireshark и tcpdump, в результате чего было создано 9 ГБ данных. |
| Типы атак | ftp\_write, guess\_passwd, Imap, ipsweep, Land, loadmodule |

|  |  |
| --- | --- |
| Название | CDMC2010 Malware API Sequence Dataset |
| Ссылка | https://dibbs.ai.arizona.edu/dibbs/csdmc2010/readme-CSMDC.txt |
| Описание | Информация о вредоносных программах, представленных в наборе данных, устарела и может не содержать слишком много информации для обнаружения сегодняшнего вредоносного ПО. Однако это может быть полезно для исторической справки или других целей.  Этот набор данных состоит из набора файлов трассировки Windows API / System-Call, предназначенных для тестирования на классификаторах, обрабатывающих последовательности. Вызовы вредоносного ПО классифицируются и помечаются как «1», а вызовы безопасного программного обеспечения - как «0». |
| Кол-во атак | Вызовы представлены в последовательном порядке. CSDM\_API\_Train.csv содержит 388 журналов. CSDM\_API\_TestData.csv содержит 378 неклассифицированных журналов. CSDM\_API\_TestLable.csv содержит классификации для CSDM\_API\_TestData.csv. |
| Типы атак | Imap, ipsweep, Land, loadmodule, multihop, Neptune, Perl, Phf, Pod, postweep |

|  |  |
| --- | --- |
| Название | Malware Training Sets |
| Ссылка | https://marcoramilli.blogspot.com/2016/12/malware-training-sets-machine-learning.html |
| Описание | Для создания набора данных использовались как статический, так и динамический анализ образцов вредоносных программ. Каждый пример представляет собой отдельный файл JSON, в котором функции и их значения перечислены в разделе свойств. Каждая функция представлена как объединение категории (например, «подпись») и действия (например, «копирует себя»). Значения закодированы в виде хэшей (чтобы ускорить обучение), и каждый хеш представляет собой наблюдение за функцией в примере. |
| Кол-во атак | APT1 292 Samples  Crypto 2024 Samples  Locker 434 Samples  Zeus 2014 Samples |
| Типы атак | Back, buffer\_overflow, ftp\_write, guess\_passwd |

**3. Классификацию атак таблицей**

|  |  |
| --- | --- |
| Отказ в обслуживании | Back, Land, Neptune, Pod, Smurf, Teardrop, Mailbomb, Processtable, Updstorm, Apache2, Worm |
| Сетевое сканирование | Satan, IPsweep, Nmap, Portsweep, Mscan |
| Получение удалённого доступа | Guess\_password, Ftp\_write, Imap, Phf, Multihop, WarezmasterXlock, Xsnoopm Snmpguess, Httptunnel, Sendmail, Named |
| Повышение привелегий | Buffer\_overload, Loadmodule Rootkit, Perl, Sqlattack, Xterm, Ps |

Отказ в обслуживании или «DoS атаки» являются одним из видов сетевых атак, предназначены для того, чтобы «затопить» целевые сети или машины большим количеством бесполезного трафика, так чтобы перегрузить атакуемую машину и в конечном итоге вывести её из строя. Основная суть DoS атаки, сделать службы, работающие на целевой машине (например, веб-сайт, DNS сервер и пр.) временно недоступными для предполагаемых пользователей. DDoS атаки, как правило, осуществляются на веб-сервера, на которых находятся жизненно важные услуги, такие как банковские сервисы, электронная коммерции, обработка персональных данных, кредитных карт.

Сетевое сканирование: злоумышленники сканируют устройства в сети, чтобы собрать сетевую информацию об этих устройствах, прежде чем запускать изощренные атаки для подрыва безопасности сети. Обычно используемые методы сканирования для сбора информации о компьютерной сети включают сканирование IP-адресов, сканирование портов и сканирование версий.

Получение удалённого доступа: попав внутрь системы, злоумышленник может загрузить вредоносное ПО, скопировать все конфиденциальные данные и использовать скомпрометированную систему для атаки на другие компьютеры или сеть в той же среде. Вредоносная программа будет продолжать красть данные даже после выхода злоумышленника из системы и может оставаться незамеченной в течение длительного периода времени.

Повышение привелегий: использование компьютерного бага, уязвимостей, ошибки в конфигурации операционной системы или программного обеспечения с целью повышения уровня доступа к вычислительным ресурсам, которые обычно защищены от пользователя. В итоге, приложение, обладающее бо́льшими полномочиями, чем предполагалось системным администратором, может совершать неавторизированные действия. «Повышением привилегий» называют ситуацию, когда пользователь компьютерной системы каким-либо образом повышает свои полномочия в этой системе (другими словами: получил возможность делать то, чего прежде делать не мог).

Для решения задачи удалённой сетевой атакой будем называть информационное разрушающее воздействие на распределённую компьютерную сеть, осуществляемое программно по доступным каналам связи.

Конкретные разновидности сетевых атак представлены в базе данных (БД) **CSE-CIC-IDS2018**.

DoS атаки — это сетевые атаки, направленные на возникновение ситуации, когда на атакуемой системе происходит отказ в обслуживании. Данные атаки характеризуются генерацией большого объема трафика, что приводит к перегрузке и блокированию сервера. ВыделяютшестьDoSатак: back, land, neptune, pod, smurf, teardrop.

U2R атаки предполагают получение зарегистрированным пользователем привилегий локального суперпользователя (сетевого администратора). Выделяют четыре типа U2R атак: buffer\_overflow, loadmodule, perl, rootkit.

R2L атаки характеризуются получением доступа незарегистрированного пользователя к компьютеру со стороны удаленного компьютера. Выделяют восемь типов R2L атак: ftp\_write, guess\_passwd, imap,multihop, phf, spy, warezclient, warezmaster.

Probe атаки заключаются в сканировании сетевых портов с целью получения конфиденциальной информации. Выделяют четыре типа Probe атак: ipsweep, nmap, portsweep, satan. Согласно источнику для обнаружения и классификации 9 из 22 типов атак достаточно 29 параметров, характеризующих сетевые соединения.

**4. Признаки таблицей**

* Анализ сетевого трафика
* Протоколы
* Действия
* Временные
* Процессы
* Изменение карты памяти и активность использования процессами
* Загрузка CPU, какими процессами

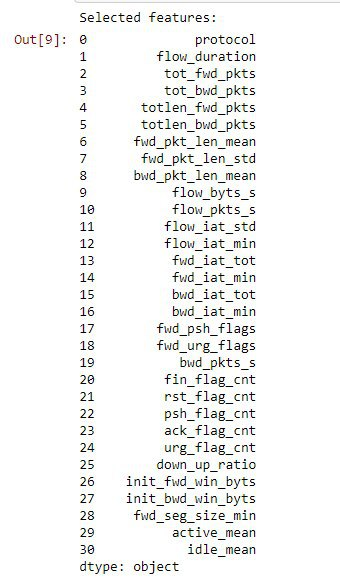
- обращение к реестру

- множественные обращения к диску

- переименования

- дописывания файлов

- системная папка



**5. Структура нейронной сети**

Совместное использование два основных вида реализации систем обнаружения вторжений на базе нейронных сетей:

* IDS на основе комбинации экспертной системы ИНС и ИНС свободно обучаемой.

Сверточная нейронная сеть

В качестве нейросетевого детектора для обнаружения сетевых атак выбрана многослойная нейронная сеть с входным слоем, одним скрытым слоем, состоящим из нейронов Кохонена, и выходным слоем. Для обучения нейросетевого детектора используется обучающая выборка, состоящая из 80 % соединений, относящихся к сетевым атакам, и 20 % соединений, относящихся к нормальным соединениям. Также, представлено применение метода многофакторного анализа (ANOVA) для сокращения размера данных для анализа сетевого трафика с целью выявления сетевых атак. Применение ANOVA позволило повысить качество обнаружения сетевых атак, а также повысить быстродействие системы за счет сокращения анализируемых данных.

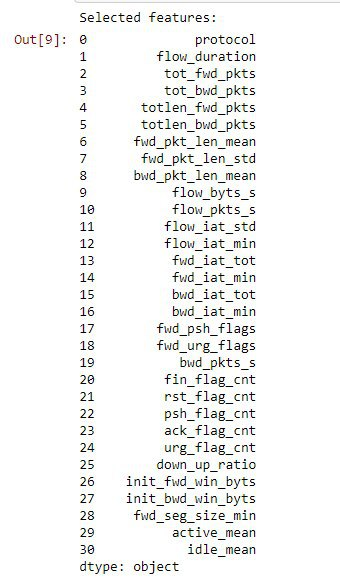
Достоинствами являются:

* более высокая скорость реакции по отношению к первому виду, поскольку присутствует всего одна ступень анализа. Скорость обработки информации в нейросети достаточна для реагирования в реальном времени на проводимые атаки до появления в системе неустранимых последствий;
* возможность для нейронной сети полностью использовать потенциал «самообучения» и выходить за пределы базы знаний экспертной системы;
* отсутствие потребности регулярного обновления базы знаний как для систем, основанных на правилах;
* нейросеть в отличие от ЭС проводит анализ информации и предоставляет возможность оценить, что данные согласуются с характеристиками, которые она научена распознавать;
* нейросеть способна анализировать данные от сети, даже если эти данные являются неполными или искаженными;
* возможность проводить анализ данных от большого количества источников в нелинейном режиме (что важно, например, при скоординированных атаках);
* способность «изучать» характеристики умышленных атак и идентифицировать элементы (например, трафик), которые не похожи на те, что наблюдались в сети прежде.

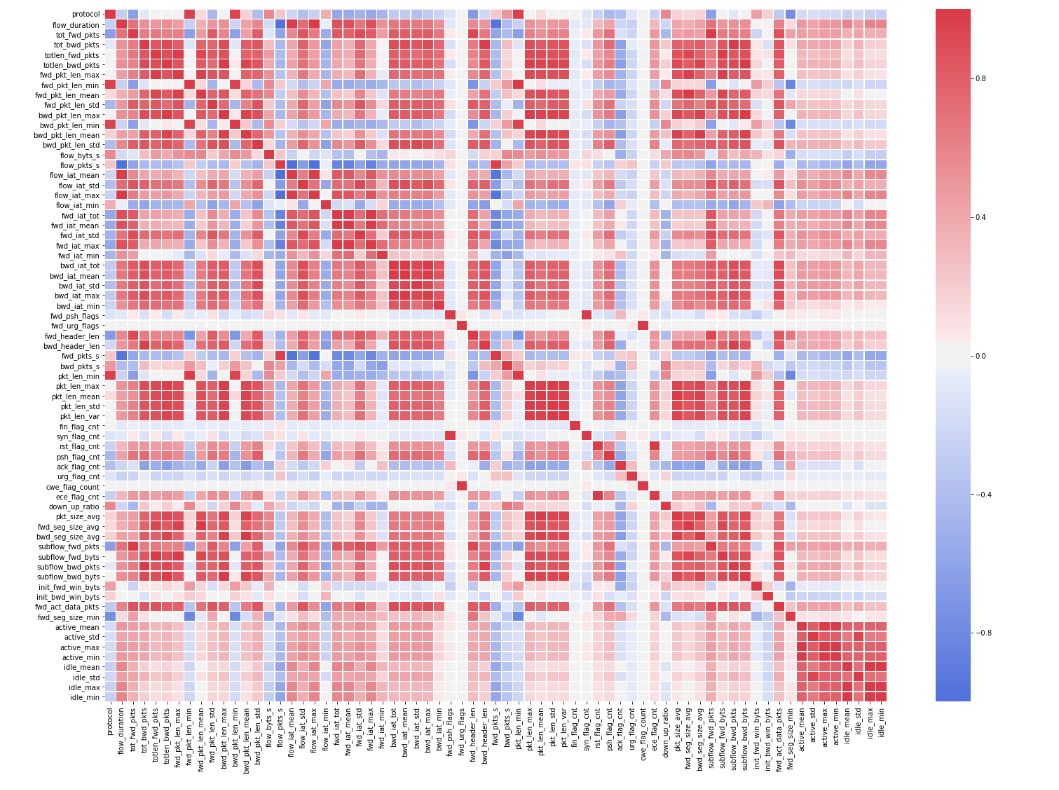
Формирование обучающей выборки в качестве ИНС использовалась структура нейронной сети Кохонена с одним входным слоем, одним скрытым слоем, состоящим из нейронов Кохонена, и выходным слоем. В качестве входных данных для нейросетевых детекторов использовались данные, взятые из базы данных **CSE-CIC-IDS2018**

Для обучения предложенного ИНС используется обучающая выборка, состоящая из 80 % соединений одного из типов атак и 20 % нормального соединения, т. е. соотношение атак к нормальным соединениям равняется четыре к одному

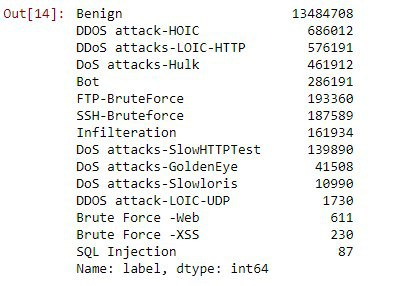
**Критерии**



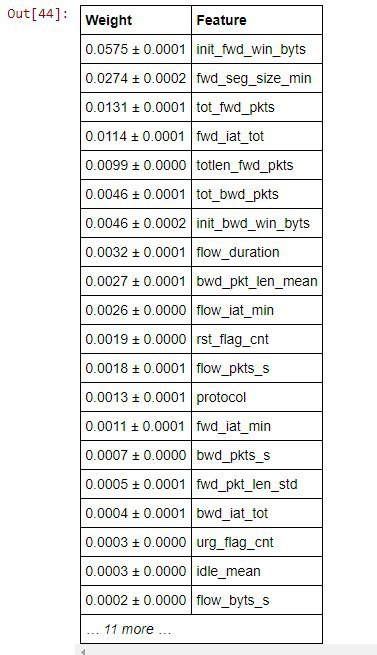
**Карта**



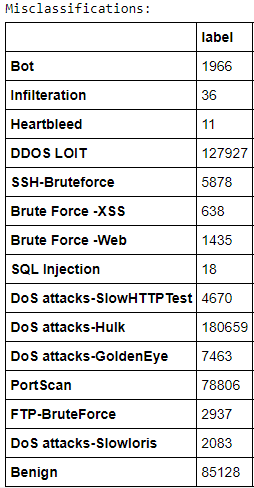
**Атаки на обучение**



**Итоговые веса**



**Какие результаты будем проверять на нашей сети с файла CSE-CIC-IDS2018**



**6. Экономическое обоснование кому продавать и как**

Клиенты: корпорации (оборот от 15ккк руб.) и предприятия среднего масштаба (оборот до 5 ккк руб)

Продукт представляется в 2х исполнениях:

* Компактный, подключаемый сервер производит аудит сети. Аренда сервера.
* Полноценный локальный, развертывается на серверах корпорации. Обслуживание по ежегодной подписке.

**7. Язык, на котором пишем прототип**

Python

TensorFlow - фреймворк нейросети

Спецификации Keras API используем в TensorFlow

**8. Протокол и уведомления об атаках**

Закрытый протокол + API для интеграции индикации и сообщений

**9. Возможность использования протоколов вроде snmp для сигнализации и встраивания в корпсети**

**10. Нейросеть пробуем проверяем сходимость пока что**

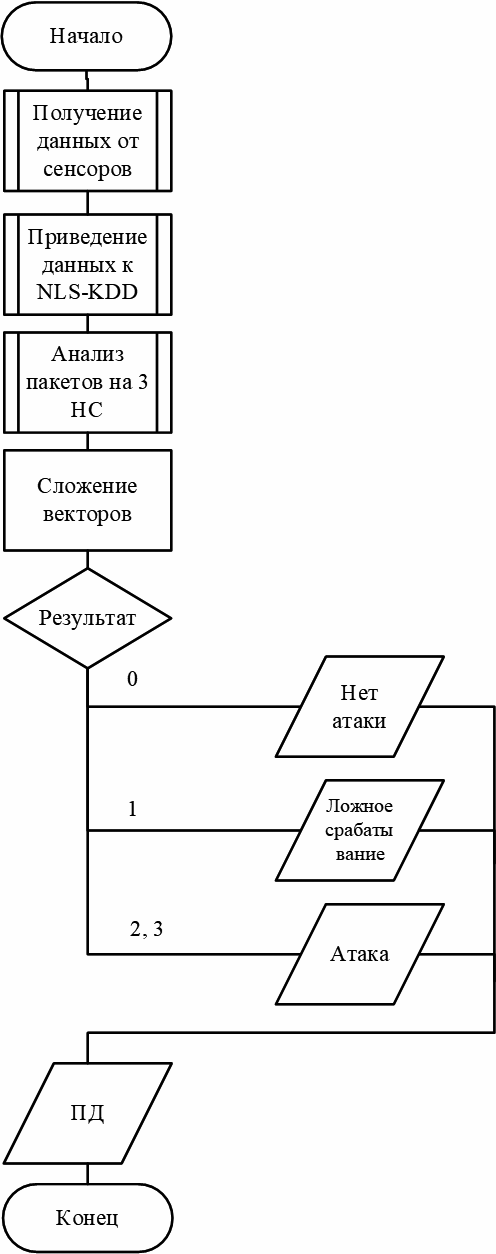
**11. Перспектива**

Для достижения поставленной в работе цели предлагается использовать три ИНС с различными топологиями: многослойный персептрон (MLP), сеть радикально-базисных функций (RBF) и самоорганизующаяся карта Кохонена (SOM). На вход каждой ИНС подается один и тот же трафик для анализа заголовков IP-пакетов совместно с неточным поиском сигнатур вторжений непосредственно в прикладном содержимом IР-пакетов. Для обучения ИНС необходимо использовать как пакеты, содержащие известные сигнатуры, так и нормальные IP-дейтаграммы [8]. В процессе обнаружения каждая ИНС определяет, насколько сетевой пакет, поданный на вход, соответствует нормальной или аномальной ситуации.

(Как делают - На выходе каждой ИНС результаты представлены в виде многомерного вектора, каждая координата которого соответствует одному из видов атак. Если ИНС детектирует атаку, то соответствующая координата принимает единичное значение, в противном случае – нулевое.) **Я бы заменил на вероятности и композитное значение.**

Алгоритм работы IDS, осуществляющей обработку и фильтрацию большого объема сетевого трафика на основе ИНС, будет состоять из следующих этапов:

1. Получение подозрительного сетевого трафика от Агента.
2. Приведение сетевых пакетов к формату NLS-KDD.
3. Анализ порции из n сетевых пакетов с помощью трех нейросетей с различной топологией MLP, SVM и SOM (количество пакетов в одной порции определяется в зависимости от характеристик сети и ИС в целом).
4. Результаты анализа заносятся в многомерные векторы.
5. Результирующий вектор получается путем сложения трех выходных векторов MLP, SVM и SOV. После этого он передается в подсистему представления данных.

[](http://samag.ru/uploads/articles/2018/09/80_83_Application_Artificial_Neural_Networks_Intrusion_Detection_Systems/Picture_3_Algorithm_of_IDS_with_INS.gif)

**Алгоритм выявления сетевых атак с использованием трех ИНС с разными топологиями**

Возможности предлагаемого алгоритма могут быть расширены как за счет увеличения количества используемых ИНС, так и применения других типов ИНС в зависимости от поставленной задачи.