# ТЕХНИЧЕСКИ УНИВЕРСИТЕТ – СОФИЯ

**ФАКУЛТЕТ ПО ТЕЛЕКОМУНИКАЦИИ**

Катедра "КОМУНИКАЦИОННИ МРЕЖИ"

**маг. инж. Каруна Сурякант Бозале**


# Изследване на алгоритми и методи за защита на информация и ресурси в комуникационни мрежи

# АВТОРЕФЕРАТ

**на дисертация за присъждане на образователна и научна степен**

**„ДОКТОР“**

Област на висше образование: 5. Технически науки

Професионално направление: 5.3 Комуникационна и компютърна техника

Научна специалност: Комуникационни мрежи и системи

**Научни ръководители:**

проф. д-р Георги Любенов Илиев

доц. д-р Мария Вълчева Ненова

**София, 2019**

Дисертационният труд е обсъден и насочен за защита от Катедрения съвет на катедра „Комуникационни Мрежи“ към Факултет по Телекомуникации при Технически Университет – София на редовно заседание, проведено на 17.06.2019 г.

Публичната защита на дисертационния труд ще се състои на 04.11.2019 г. (понеделник) от 15.00 часа в Конферентната зала на БИЦ на Технически университет – София на открито заседание на научното жури, определено със заповед № ОЖ-5.3-71 от 15.07.2019 г. на Ректора на ТУ-София в състав:

1. Проф. д-р инж. Владимир Костадинов Пулков – председател
2. Проф. д-р инж. Георги Любенов Илиев – научен секретар
3. Доц. д-р инж. Филип Атанасов Цветанов
4. Доц. д-р инж. Станимир Михайлов Садинов
5. Доц. д-р инж. Иван Стефанов Узунов

Рецензенти:

1. Проф. д-р Владимир Костадинов Пулков
2. Доц. д-р Филип Атанасов Цветанов

Материалите по защитата са на разположение на интересуващите се в кабинет 1254, блок 1, Технически Университет – София.

Дисертантът е задочен докторант към катедра „Комуникационни мрежи“, Факултет по Телекомуникации. Изследванията по дисертационният труд са направени от автора, като резултатите от тях са публикувани.

Автор: маг. инж. Каруна Сурякант Бозале

Заглавие: Изследване на алгоритми и методи за защита на информация и ресурси в комуникационни мрежи

Тираж: 30 броя

Отпечатано в ИПК на Технически университет – София

# I. ОБЩА ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД

## Актуалност на проблема

Използването на Интернет в наши дни като основна среда за предаване и обмен на данни е изключително актуално. Крайните потребители имащи достъп до глобалната мрежа, обаче са уязвими към най-различни видове злонамерени атаки. Има много разработки за с цел решаването на този проблем. Те могат да бъдат групирани въз основа на определени характеристики. Основната им цел е откриването на атаката още в самото начало и възможността да се предотврати компрометирането на комуникацията. Главното предизвикателство пред разработването на такива решения е да се определи прецизността на откриване, тъй като повечето от настоящите подходи се оценяват чрез използването на публично достъпни масиви от трафични данни, а не чрез мрежов трафик в реално време. Съществуващите методи за анализ срещат проблеми, когато се работи с данни в реално време, защото те съдържат и съществено количество шум, което би могло да доведе до грешно прогнозиране. Това води до необходимостта да се проектират нови системи за откриване на интрузии в комуникационни мрежи, с цел разпознаване на различни видове заплахи за сигурността, както в реално време, така и в офлайн потоци от данни. За целта може да се приложи, например предварителна обработка на данните, избор на характеристики и алгоритми за класификация на данните. Смята се, че точността и ефективността на откриване на атаки се основава на коректността на прихванатия набор от данни. Актуалността на настоящия труд е в използването на различни алгоритми и методи и техни модификации за разрабоване на система за откриване на атаки. Предложеното решение може да се определи като модифицирана система за откриване на интрузии, базирана на Naive Bayes подхода.

## Цел на дисертацията и основни задачи

Основаната цел на дисертационния труд е: Формулиране и разработване на нови алгоритми и методи за предварителна обработка на данни, избор на характеристики и класификация на данни. С предложените алгоритми и методи да се проектира и оцени поведението на нова система за детекция на интрузии, да се подобрят съществуващите класификатори, за да се повиши ефективността на откриването на атаките. За да се постигне тази цел, трябва да се изпълнят следните задачи:

1. Да се проучат основните заплахи и предизвикателства за сигурността в комуникационните мрежи, като се обърне основно внимание на атаки от типа DOS, DDOS, R2L и U2R.
2. Да се акцентира върху анализа на характеристиките на най-новите атаки на приложния слой, от тип DOS. Да се изследва въздействието на този вид атака, както и нейните специфични характеристики и механизми за защита.
3. Тъй като DDOS атаките се увеличават по видове и обхват, като по този начин се повишава сложността, е необходимо усилията да се фокусират върху откриването на DDOS атаки на приложния слой от OSI модела. Трябва да се предложи модел за DDOS атака на приложния слой, разработен с помощта на UML моделиране.
4. Да се предложат техники за филтриране на данни с цел подобряване на анализа им и намаляване на неговото времетраене. Да се реализира прихващането на данните в реално време. Тъй като нивото на шума, дължащо се на неподходящи масиви от данни и грешки при събирането им е високо, да се използват техники за предварителна обработка на данните.
5. Да се предложи нов метод за извличане на характеристики, който да подобри точността на класификация на събраните данни. Да се извърши оценка на ефективността на предложения метод, като се сравни с различни видове класификатори като KNN, ANN, SVM и NB.
6. Да се подобри поведението на системата за откриване на интрузии и ефективността на класификацията като се запази ниска изчислителната сложност.

**Научна новост**

Научната новост в дисертационния труд е свързана с предлагането на нови техники за подобряване на сигурността в Интернет. Основният резултат е нова система за откриване на интрузии MNBIDS, която има три основни характеристики: предварителна обработка на данните, избор на подходящи характеристики и подобрен метод за класификация. Системата MNBIDS е сравнена със съвременните IDS подходи. Новостите в системата могат да се обобщят по следния начин. Naive Bayes методът е модифициран MNB, за да се намали изчислителната му сложност, така че да може ефективно да се прилага с прости оператори в процеса на откриване и класифициране. При обучението се прилага MNB за настройка на параметрите в различните класове данни, което позволява подобряване на ефективността на класификацията. Резултатите

от MNB се сравняват със съществуващите класификатори и се доказва, че са по-добри от останалите IDS системи. Предложеният модел на модифицирана Naive Bayes система за откриване на интрузии MNBIDS е сравнен от гледна точка на прецизност, чувствителност и точност със сходни IDS системи. Получените резултати показват, че системата MNBIDS постига подобрение на точността с приблизително 17% в сравнение с другите използвани техники.

**Практическа приложомост**

Всички разработени алгоритми и методи, както и предложената модифицирана Naive Bayes система за откриване на интрузии MNBIDS са изследвани и анализирани чрез симулационни експерименти. MNBIDS системата може да намери практическо приложение при интегрирането ѝ в комплексни решения за защита на информация и ресурси в комуникационни мрежи.

**Публикуване на резултатите от изследванията в дисертацията**

Проучванията и изследванията съдържащи се в дисертационния труд са проведени в периода 2016 – 2019. Те са представени в 9 авторски публикации, 7 от които на международни конференции и 2 в научни списания. Една от публикациите е самостоятелна, а в останалите 8 са в съавторство. Международните конференции са: Global Wireless Summit (GWS) 2016; International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS) 2017; International Conference on Smart City and Emerging Technologies 2018; International Conference on Computer Networks and Inventive Communications Technologies (ICCNCT) 2018; International Conference on Computational Techniques, Electronics & Mechanical Systems 2018; International Conference On Advanced Technologies for Societal Applications 2018. Научните списания са: International Journal of Academic Research and Development 2017; International Journal of Advanced Research and Development 2017.

**Структура и обем на дисертационния труд**

Дисертационният труд е написан на английски език и е в размер на 146 страници, съдържа 4 глави, заключение с посочени основни приноси, списък на фигурите, списък на таблиците, списък на използваните съкращения, списък на публикациите на автора, списък на използваната литература. В работата има 52 фигури и 5 таблици. Използвани са 169 литературни източника, всичките на латиница като 75% от тях са от последните 10 години. Номерата на фигурите и таблиците в автореферата отговарят на тези в дисертационния труд.

# II. СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯ ТРУД

# ГЛАВА 1. КРИТИЧЕН АНАЛИЗ НА РАЗЛИЧНИТЕ ТИПОВЕ ЗАПЛАХИ ЗА СИГУРНОСТТА И DDOS АТАКИТЕ

## 1.1. Литературен обзор на различните типове заплахи и възможните решения за преодоляването им

Заплахите към сигурността на комуникационните мрежи са важен и широко изследван проблем. В дисертационния труд са представени научни публикации разглеждащи проблема и възможни решения в зависимост от конкретния случай. В [13], е представено решение фокусирано върху разпределена атака за отказ от услуга DDOS. Представени са специфичните характеристики и етапи на протичането на DOS и DDOS атаки. Това са най-често прилаганите подходи за атака, които оказват влияние върху сигурността и спецификата на администрирането реално време [14]. Разгледани са съществуващи DOS атаки и нови подходи за защита в Интернет и системи с отдалечен достъп. Акцентът им е върху различни стратегии за осъществяване на самата атака с цел разбирането на тяхната специфика. Наборът от характеристики на DOS са описани подробно в [18]. Авторите показват използването и анализа на различни типове атаки - Ping of Death, TCP SYN Flood, и разпределена DOS (която е и най-тежката). Атаката Ping of Death е реализирана срещу Microsoft Windows 95 система. TCP SYN Flood атаката може да се скрие зад Microsoft Windows 2000 IIS FTP Server. Атаката от тип „Circulated DOS" може да се реализира чрез пресъздаване на „зомби" програма с вид на „Ping of Death" атака [19]. Авторите представят потенциалните последствия, които може да нанесат DOS атаките и как те да се неутрализират, да се възстановят ресурсите и намалят последиците от тях. Те използват различни класификатори, като най-често това е Naive Bayes, полиномни Naive Bayes, многослойни персепторни, RBF, Random Forest и други. След това са разгледани в детайли поведението на най-тежките DOS атаки - разпределените DOS [21]. Представени са и последните изследвания на заплахи в базираните на облачни изчисления комуникационни мрежи. DOS и DDOS атаките са риск не само за администрирането в облака. Това е така поради голямото многообразие на нови открити слабости в облака [22]. Разгледани са и особеностите на DOS и DDOS атаките при облачните изчисления и по-специално XML-DOS и HTTP-DOS , както и структури за осигуряване на защита [A1].

DDOS атаката се разпространява и координира чрез няколко компрометирани хоста. Зловредният трафик, генериран и изпратен от компрометираната система, е много подобен на нормалния трафик, което затруднява откриването и предотвратяването им. Атаките към честотната лента са най-типичните DDOS атаки. Най-популярните атаки включват наводнения от ICMP, UDP, HTTP трафик и други с фалшиви пакети. Понастоящем нападателите използват не само конвенционални

методологии за атака, но и наемат центрове за данни на доставчици на облачни услуги, за да стартират атаки с изчерпване на ресурс чрез отдаване под наем или компрометиране на облачни системи.

## 1.2. Обзор на подходи за реализация на сигурни комуникации

В тази част на дисертационния труд е направен обзор на състоянието на проблема и са показани някой от най-новите решения за постигане на сигурност в комуникационни мрежи чрез използване на различни подходи за реализация на системи за откриване на интрузии. Вниманието е насочено към DDOS атаките, тъй като те могат да предизвикат огромни проблеми от гледна точка на сигурността и да компрометират комуникацията между легитимните потребители на услуги [A2]. През годините се предлагат решения използващи различни видове класификатори [23]. Някои от получените експериментални резултати показват висока степен на откриване на атаките, което е много обещаващо [24].

Поради основните характеристики на LSSVM-IDS системата за откриване на интрузии тя постига по-голяма прецизност и по-ниски изчислителни разходи, което я прави една от най-добрите в нейния клас [26]. Системата IIDPS работи с профили на легитимните потребители, като следи начина на използване на ресурсите и въз основа на това решава дали даден потребител е легитимен или не [27]. Друг подход за откриване на атаки и разграничаване на легитимен от нелигитимен трафик е използването на статистически модели като ARIMA, чрез които да се следи за нехарактерни трафични модели [29].

## 1.3 Заключение

В първа глава са анализирани основните проблеми за сигурността в комуникационните мрежи. На различни нива в тях могат да бъдат открити уязвими места за реализирането на атаки. Проучени са и анализирани различни заплахи към сигурността - атаки на приложения, ASYN, пакетна фрагментация, смъртоносен пинг и наводняване. В литературния обзор са проучени и описани голям брой научни разработки на изследователи по темата. Сравнени са основните видове решения, като класификатори за откриване на DDOS, филтриране в реално време на потоци от данни (RTFF), IDS базирани на LS-SVM, IIDPS, ARIMA и алгоритми за „deep learning”. В заключение са разгледани някои от недостатъците на тези решения и възможностите за тяхното подобрение с цел повишаване на нивото на сигурност и защита на информацията и ресурсите в комуникационните мрежи.

## 1.4 Приноси към първа глава

1. На база на проведения критичен анализ в първа глава са определени основните проблеми свързани със защитата на информацията и ресурсите в комуникационните мрежи. Става ясно, че съществуващите IDS решения не

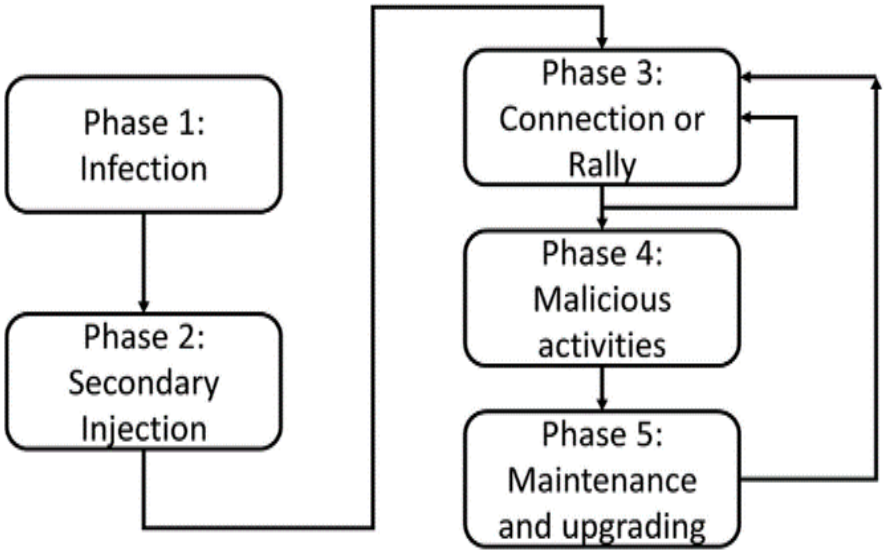
могат да осигурят изцяло мрежовата сигурност. Съществуващите методи не са толкова ефективни, когато се работи с големи обеми от данни.

2. Възможен подход за решение на проблема е разработването на нови методи за системи за детекция на интрузии и модифицирането на съществуващи класификатори за да се осигури по-ефективна работа. Това е свързано с разработването на алгоритми за предварителна обработка на данните и избор на подходящи характеристики.

# ГЛАВА 2. АНАЛИЗ НА ВЛИЯНИЕТО НА БОТНЕТ ВЪРХУ МРЕЖОВАТА СИГУРНОСТ И ТЕХНИКИ ЗА ЗАЩИТА ОТ DDOS АТАКИ НА ПРИЛОЖНО НИВО

## 2.1. Жизнен цикъл и архитектура на БОТНЕТ

Както се вижда от Фиг.2.2. жизненият цикъл на Ботнет може да се раздели на пет основни фази – заразяване, проникване, свързване, злонамерени действия и поддръжка.




**Фигура 2.2. Жизнен цикъл на Ботнет**

Инсталирането на злонамерени програми, осигуряващи механизми за отдалечен контрол на компютри се наричат „ботове". Когато тези ботове са контролират в рамките на една и съща структура и водят до сериозни заплахи към Интернет като система, се наричат Ботнет [A3]. Откриването и премахването им е задача пред компютърните специалисти за да се осигури сигурността в цифровото пространство. Най-често ботовете се използват за изпращането на спам, атаки като DDOS и други, които са описани подробно в [33] и [34]. Предложена е методология за анализ и отслабване на ефекта от атаката. Има детайлно описание на иницирането и разпространението на „Storm Worm" ботове. Това проучване позволява да се оцени количеството на заразените машини, да се анализират и инфилтрират подробности

относно тези ботнети. Учените препоръчват два основни метода чрез които комуникацията между атакуващият и заразените машини се прекъсва и спира. Те изследват ефективността на методите в [37] и [38].

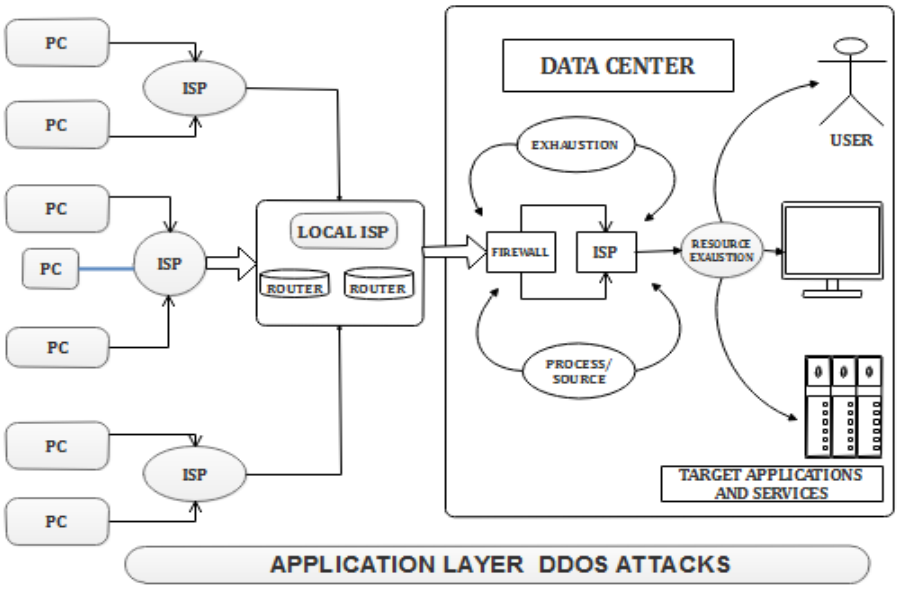
## 2.2. DDOS атаки на приложно ниво и методи за тяхното откриване

При анализирането на характеристиките на мрежовия трафик в последно време се въведе понятието „flash crowd". В мрежата с този термин се описва състоянието, когато много голям брой потребители едновременно имат достъп до популярен уебсайт. Това води до огромен трафик към уебсайта и като резултат той не може да се достъпи. Тъй като трафикът и големият обем са основни характеристики на App-DDOS атаките или „flash crowd", използваните в момента техники е трудно да ги разграничат само по статистическите характеристики на трафика. Ето защо App-DDOS атаките могат да са неуловими и с по-голяма опасност за търговски уебсайтове, отколкото класическите DDOS атаки. Това е едно от основанията за разработването на нови решения за защита срещу такъв тип DOS атаки.

Поради отворения характер на Интернет мрежовите оператори много често отказват да внедряват нови и потенциално трудни за реализация механизми за осигуряване на определени нива на сигурност. Подходи като Secure Overlay Services (SOS) и Mayday, представляват една добра алтернатива. Това се дължи на факта, че те не изискват промени в протоколите и рутерите, което от своя страна води и до минимална нужда от влагане на ресурси от страна на Интернет доставчиците (ISPs). Реализацията на такива системи включва множество мрежови възли, които биха могли да изпълняват функциите на защитни стени на първо ниво, разделяйки трафика на – нормален и нежелан за различните видове потребители и крайни устройства. Един от най-обещаващите подходи е базиран на TCP Congestion Window за откриване на атаки реализирани чрез Ботнет. Напоследък все по-широко се прилагат и техники използващи машинно обучение [A4].

## 2.3. Откриване на DDOS атаки на приложно ниво чрез използване на UML моделиране

Фиг. 2.6. описва общия модел на DDOS атаката на приложно ниво. В мрежите на Ботнет, атакуващите компютри са географски разделени, използвайки различни интернет връзки, като по този начин е трудно да се контролират атаките. Целта е капацитета на защитната стена да е с изчерпани ресурси от потока от данни, които напълно заемат наличния ресурс. Това наводняване причинява проблеми свързани с претоварване на атакуваната мрежа или сървър. То води до сериозни смущения в обслужването на легитимните потребители, които се опитват да получат достъп.




**Фигура 2.6. UML диаграма на DDOS атака на приложно ниво**

Структурата на центъра за данни се състои от защитна стена, интернет доставчик, приложения и услуги, които са целта на атаката. Чрез претоварване на сървъра с пакети от данни се изчерпва капацитета за предоставяне на услуги на приложно ниво. Това може да създаде изключително негативни последици за бизнеса във финансовия сектор, академичните институции и индустрията, ако те са базирани на достъп до уебсайтове. В настоящият труд чрез използване на UML моделиране ясно се дефинират различните етапи на атаката и по този начин се намалява времето за нейното откриване [А5].

## 2.4. Комплексни решения за подобряване на сигурността на приложно ниво

Широкообхватно проучване на атаките на приложния слой, което осигурява и по-добро разбиране на текущото решение на проблема и възможностите за бъдещи изследвания за защита от такива атаки е представено в [127]. Това проучване е разширено и в представените разработки и анализи в [128]. Показано е, че DDOS атаките на приложно ниво спомагат за атаки базирани на наводняване на мрежата жертва. Основна цел на атаките на приложно ниво, като най-високо ниво в OSI модела е HTTP (Hypertext Transfer Protocol ) протокола [139]. Авторите изтъкват факта, че поради сложността и многообразието от мрежови протоколи и услуги, атаката DDOS на приложно ниво е водеща. В [140] е предложен нов метод за откриване на DDOS атаки на приложно ниво, базиран на One Class – Support Vector Machine (OC-SVM).

## 2.5. Заключение

Във втора глава подробно са проучени DDOS атаките на приложно ниво, техните разновидности, категории и работни механизми. Акцентът тук е върху анализа на

различните подходи за DDOS атаки и защитата от тях. В началото са описани архитектурата на DDOS атаките реализирани чрез Ботнет. След това са сравнени различни решения за повишаване на сигурността. Чрез използването на UML моделиране е създаден обобщен модел, който може да се приложи в системи за откриване на DDOS атаки на приложно ниво.

### 2.6. Приноси към втора глава

1. Анализиран е метод, базиран на TCP Congestion Window, за откриване на DDOS атаки реализирани чрез Ботнет. Анализът е направен, като е използван CUSUM алгоритъм реализиран върху платформа на мрежов симулатор. При този метод целта е да се оптимизират параметри, като най-кратко време за откриване, ниски стойности на вероятността за грешно откриване и малка изчислителна сложност.
2. Направен е анализ на метод използващ машинно обучение за откриване на DDOS атаки реализирани чрез Ботнет. При него се използват техники базирани на Identifier/Locator подхода за откриване на DDOS атаки. Едно от най-големите преимущества на този метод е трудното осъществяване на контрол от страна на атакуващия върху мрежите Ботнет.
3. Направен е анализ на характеристиките на появилите се напоследък DDOS атаки реализирани на приложния слой. Показано е влиянието, което оказват този тип атаки. Изследвани са специфичните характеристики на DDOS атаките с наводняване на мрежата и методите за защита при този тип атаки.
4. Използвайки UML моделиране е разработен модел на DDOS атака реализирана на приложния слой. UML моделът ясно дефинира различните етапи на атаката. Целта е намаляване на времето за откриване на DDOS атаки на приложно ниво.

# ГЛАВА 3. НОВИ АЛГОРИТМИ ЗА ПРЕДВАРИТЕЛНА ОБРАБОТКА НА ДАННИТЕ И ИЗБОР НА ПОДХОДЯЩИ ХАРАКТЕРИСТИКИ

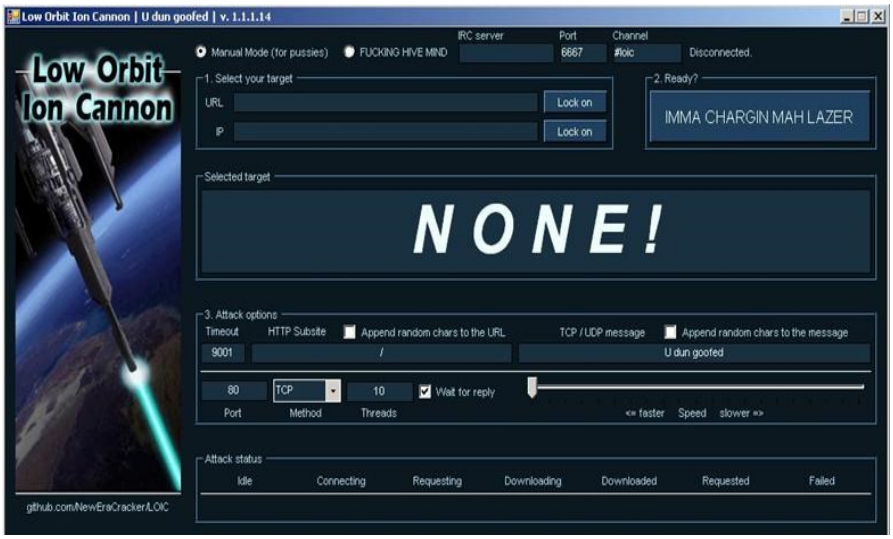
Смята се, че прецизността и ефективността при откриването на атаки се основава на коректността на прихванатия трафик от данни. Първоначално е било проблем откриването и премахването на шума в данните от логовете, което е довеждало до получаването на некоректни резултати. Поради тази причина в дисертацията изследванията започват с прихващането на трафик (нормален и нежелан) в реални компютърни мрежи. Това е използвано като база за разработването на алгоритъм за тяхната предварителна обработка, за да се отстранят нежеланите компоненти и така да се подобри способността за откриване на атаките. Разработеният алгоритъм е както за реален трафик, така и за такъв от базата с данни KDDcup. В структурата на предложеното решение се използва LOIC DDOS (Low Orbit Ion Cannon)

генератор на атаки по време на прихващането на пакетите, както и KDDcup за другите типове, като DoS, R2L, L2R, и Probe. След предварителната обработка се прилага и нов алгоритъм за избор на характеристики, за да се селектират релевантни и уникални характеристики и да се повиши точността на откриването на атаките.

Използван е и пакетен „снифър", често определян и като компютърна мрежа, която е съвкупност от хардуерни компоненти и компютри свързани чрез комуникационни канали, позволяващи, споделянето на ресурси и информация. Тук сигурността е от изключително значение. Предложеното решение е опит да се повиши способността за откриването на различни типове атаки, максимално бързо.

### 3.1. Генератор на атаки

С цел генерирането на атака по време на прихващането на пакети в реално време се използва софтуерен генератор на атаки - LOIC. Той е генератор на атаки с отворен достъп.




**Фигура 3.3. Модифицирана версия на LOIC с възможност за IRC връзка**

Описание на полетата:

- **IDLE:** Показва количеството на неактивни низове. То трябва да бъде нула за по-висока производителност на заплахата.
- **Connecting:** Показва количеството низове, които се опитват да се свържат със засегнатия сървър.
- **Requesting:** Показва броя низове, които изискват данни от засегнатия сървър.
- **Downloading: П**оказва броя низове, които започват да изтеглят данни от сървъра.

## 3.2. KDD база от данни

Освен прихващането и наблюдението на пакети в реално време, се използва и набор от оценки, за да се определи ефективността на предложените алгоритми. Данните от KDD99 са базирани на откриване на проникване, което е започнало с проекта DARPA (Defense Advanced Research Projects Agency) през 1998.

- **Denial of Service (DOS):** чрез използването на услуги нападателите се опитват да отстранят легитимните потребители;
- **Remote to Local (r2l):** на машината, която е „жертва" нападателят няма акаунт и се опитва да разшири достъпа си;
- **User to Root (u2r):** атакуващият има локален достъп до жертвата и се опитва да придобие превилигирован достъп;
- **Probe:** атакуващият се опитва да разшири обема на получаваната информация за жертвата.

Различните видове характеристики са групирани по следния начин:

- *Основни характеристики*: Основните функции могат да бъдат извлечени от заглавните полета на пакетите, без да се проверява полезния товар;
- *Характеристики на съдържанието*: Информацията за домейна се използва за да се оцени полезния товар на оригиналните TCP пакети. Това включва характеристики като броя на опитите за некоректно влизане в акаунта;
- *Времеви трафични характеристики*: Те са разработени с цел да прихващат характеристики, които са във втория времеви период. Пример за такива са броя на връзките към даден компютър от мрежата по време на втория времеви интервал;
- *Хост базирани трафични характеристики:* Използват се данни получени във времето (а не само за конкретен времеви период) за броя на връзките – в изследвания случай 100. Така характеристиките на хоста служат за оценка на заплахите, които преминават през интервали по-дълги от две секунди.

## 3.3. Нов алгоритъм за предварителна обработка на данните

В тази глава е разработен нов алгоритъм включващ предварителна обработка на набор от данни с цел подобряване на поведението и успеваемостта за откриване на нежелани събития в системите за детекция на интрузия. Данните получени по време на събирането им в реално време и от KDD Cup 99 са обработени с този алгоритъм [A6].

За целите на анализа на резултатите от алгоритъма за предварителна обработка се използват следните основни параметри:

$$\text{Точност (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (3.1)$$

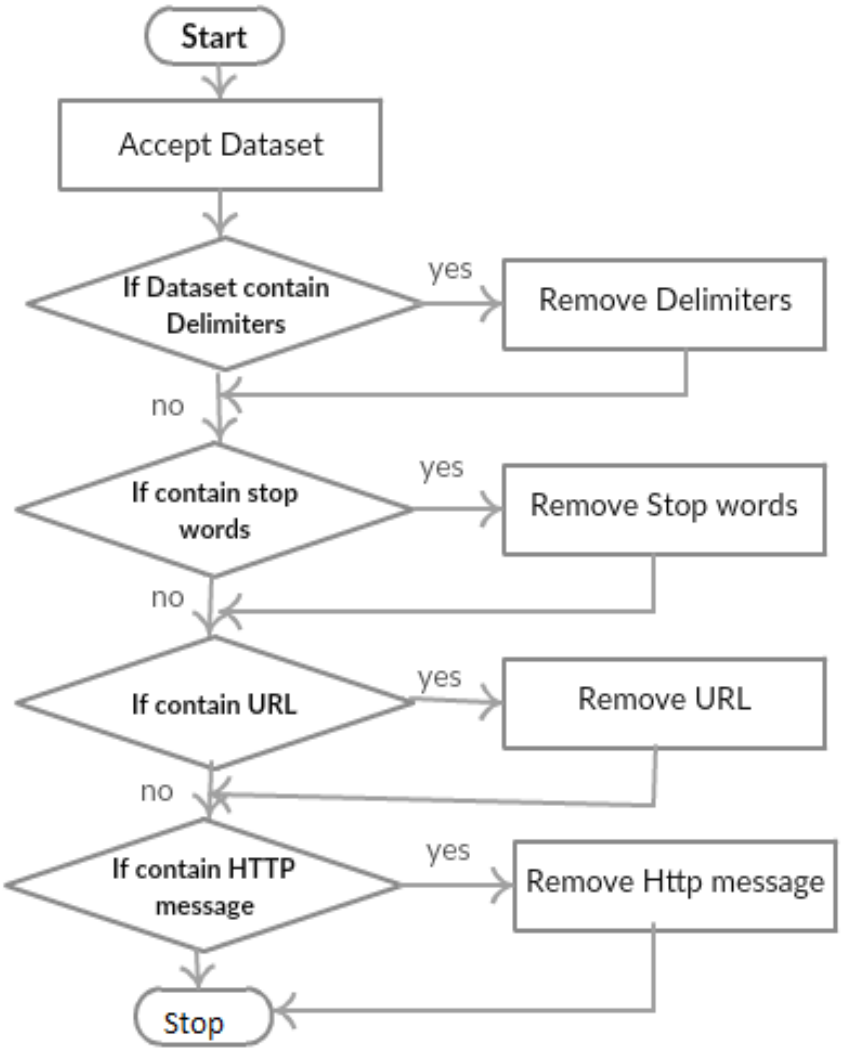
$$\text{Чувствителност (Recall)} = \frac{TP}{TP+FN} \quad (3.2)$$

$$\text{Прецизност (Precision)} = \frac{TP}{TP+FP} \quad (3.3)$$

$$\text{F-Measure} = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (3.4)$$

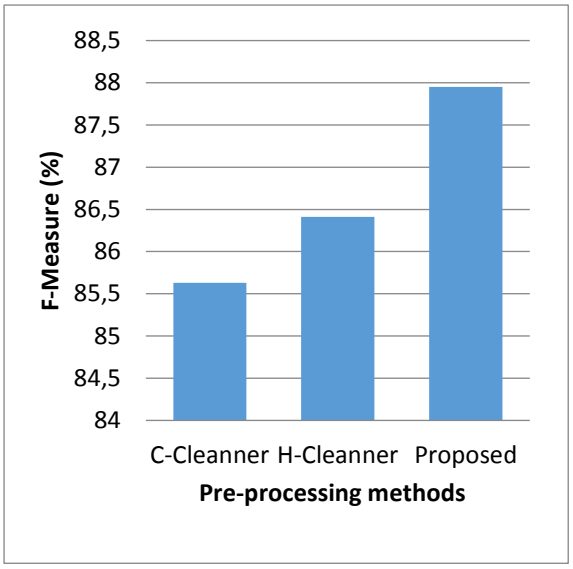
където TP - true positive, TN - true negative, FP - false positive, FN - false negative.

На Фиг. 3.5 е показана графично структурата на алгоритъма на работа и отделните фази, през които се преминава, използвайки KDD CUP 99.

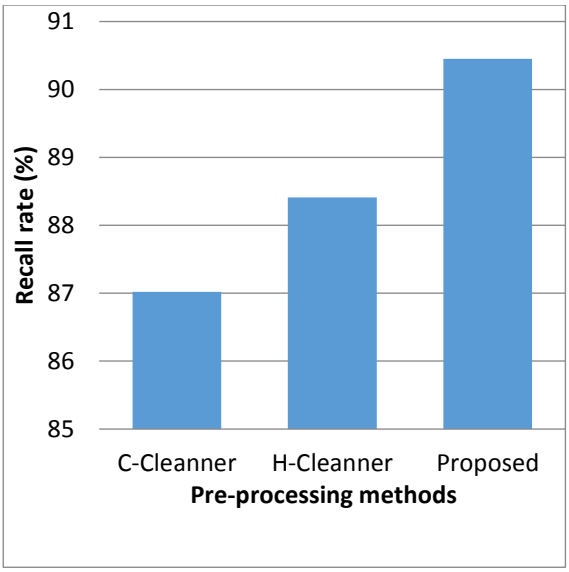



**Фигура 3.5. Диаграма на разработения алгоритъм за обработка на данните**

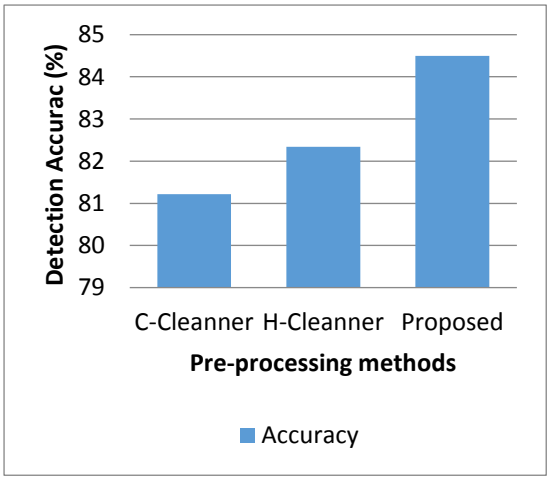
Фигури 3.11, 3.12, 3.13 и 3.14 показват сравнителни резултати със съвременни методи като H-Cleaner [156] и C-Cleaner [158].



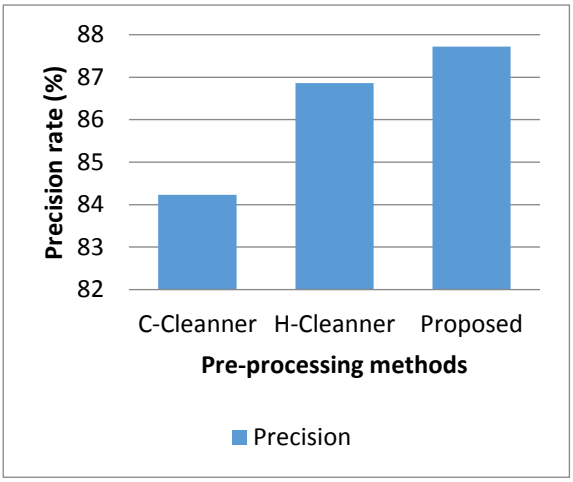

**Фигура 3.11. Анализ на параметъра F-measure**




**Фиг. 3.12. Анализ на параметъра чувствителност**




**Фигура 3.13. Анализ на параметъра точност**




**Фигура 3.14. Анализ на параметъра прецизност**

От фигурите се вижда, че предложеният алгоритъм за предварителна обработка на данните превъзхожда по всички параметри два от най-разпространените алгоритми. Това се дължи на новия подход за премахване на шумната част от данните, който се използва в предложения алгоритъм.

## 3.4. Нов хибриден алгоритъм за извличане на характеристики

Извличането на подходящи характеристики от данните е основен процес в системите за откриване на интрузии/атаки. Процесът на извличане се състои в търсене

на повтарящи се характеристики и определянето на оптималното множество от характеристики, с което да се работи. В настоящия труд е предложен хибриден алгоритъм Hybrid Feature Selection Method (HFSM), който има две основни фази [A7]. В първата фаза се извършва предварително търсене, за да се елиминират излишните и неподходящи характеристики от оригиналните данни. Във втората фаза, базирана на филтри, се запазват най-важните характеристики и се използват за конструиране на класификатори за съответните класове. Предимството на този подход е, че се идентифицират най-информативните характеристики на трафичните данни, което води до по-висока производителност.

**Algorithm 2: Hybrid Feature Selection Algorithm**

**Input:** Pre-processed data set $P=\{f_i, i=1,2....n\}$

**Output:** $F$ – the optimized features subset

1. Initialization: F=0;
2. Calculate *Corr (C; fi)* for each feature $i=1,2... ..., n$;
3. $n_f = n$; Select feature $f_i$ such that:

   $argmax (Corr (C; f_i)), i=1,2....n_f,$

   Then set $P \leftarrow P \setminus \{f_i\}; C \leftarrow C \cup \{f_i\}; n_f = n_f - 1$;

4. While ≠ 0 do

   Calculate $G_{corr}$ in to find $f_i$ where $i \in (1, 2, ... ..., n_f);$

   $n_f = n_f - 1;$

   $P \leftarrow P \{f_i\};$

   *If* $G_{corr} > 0$ then

   I $F \leftarrow F \cup \{f_i\}$.

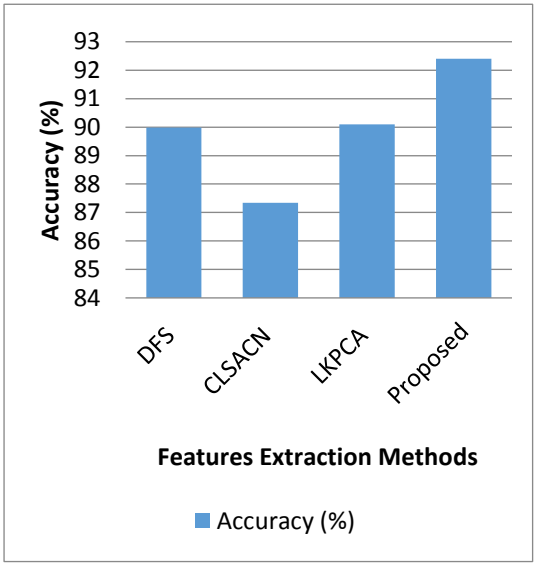
   End

End

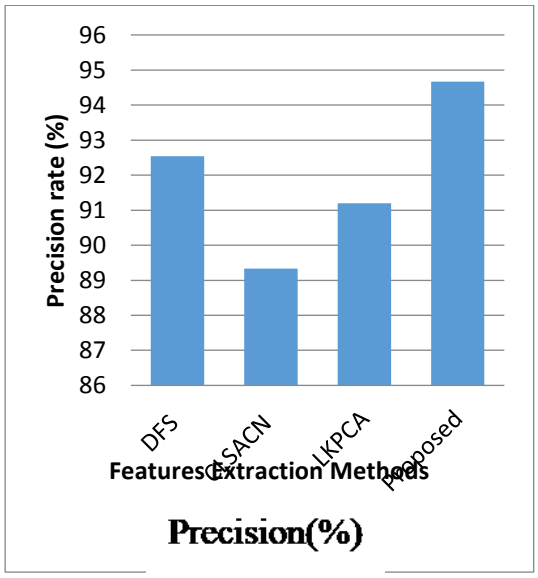
5. Sort F according to the value of $G_{corr}$ of each selected feature.
6. Return F.

## 3.5. Анализ на работата на HFSA

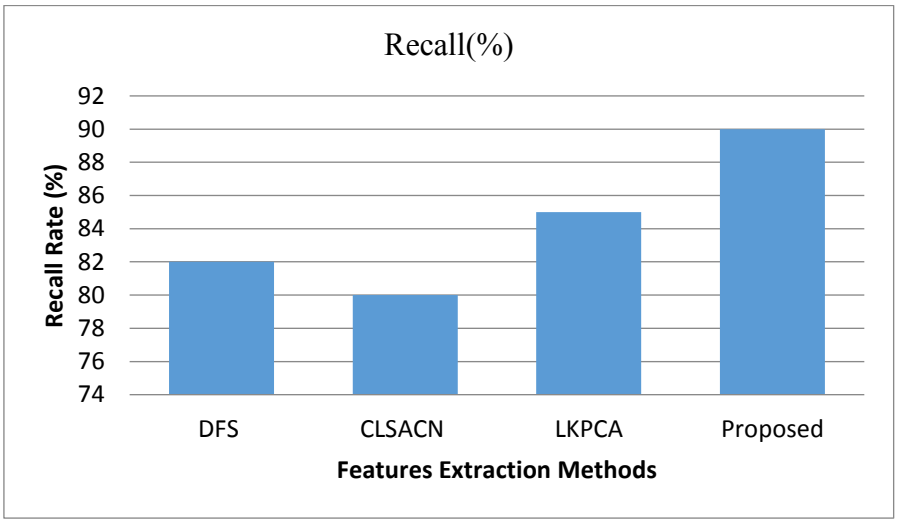
За да се оцени ефективността на предложения алгоритъм, той се прилага в комбинация с новия метод за предварителна обработка на данните разработен в настоящия труд [A7]. Двата алгоритъма - с предварителна обработка и извличане на характеристики могат да се интегрират в рамките на комплексна система за откриване на интрузии. Експериментите са върху трафик в реално време, както и върху данни от базата KDD CUP 99. Наборите от данни са разделени на тестови и обучаващи в съотношение 30% към 70%. За оценка на предложения алгоритъм се използва "Naive Bayes" класификатор.




**Фигура 3.15. Оценка на точността**




**Фигура 3.16. Оценка на прецизността**




**Фигура 3.17. Оценка на чувствителността**

Разработеният алгоритъм се оценява чрез сравнение със съвременни методи за избор на характеристики като DFS [164], CLSACN [165] и LKPCA [166]. От Фиг. 3.15, 3.16 и 3.17 се вижда, че той ги превъзхожда по отношение  и на трите параметъра използвани за оценка на поведението му - точност, прецизност и чувствителност.

### 3.6. Заключение

В трета глава са представени два от основните приноси в дисертационния труд. Те се състоят в разработването на нови алгоритми за предварителна обработка на данни и подбор на подходящи характеристики за класифициране на данните. Поведението на новите алгоритми е оценено от гледна точка на точност, чувствителност и прецизност. Експериментите са извършени, както върху прихванати в реално време данни, така и върху данни от бази за сравнение за да се докаже скалируемостта и ефикасността на предложените алгоритми. Двата метода могат да се комбинират и интегрират в комплексна система за откриване на интрузии/атаки.

### 3.7. Приноси към трета глава

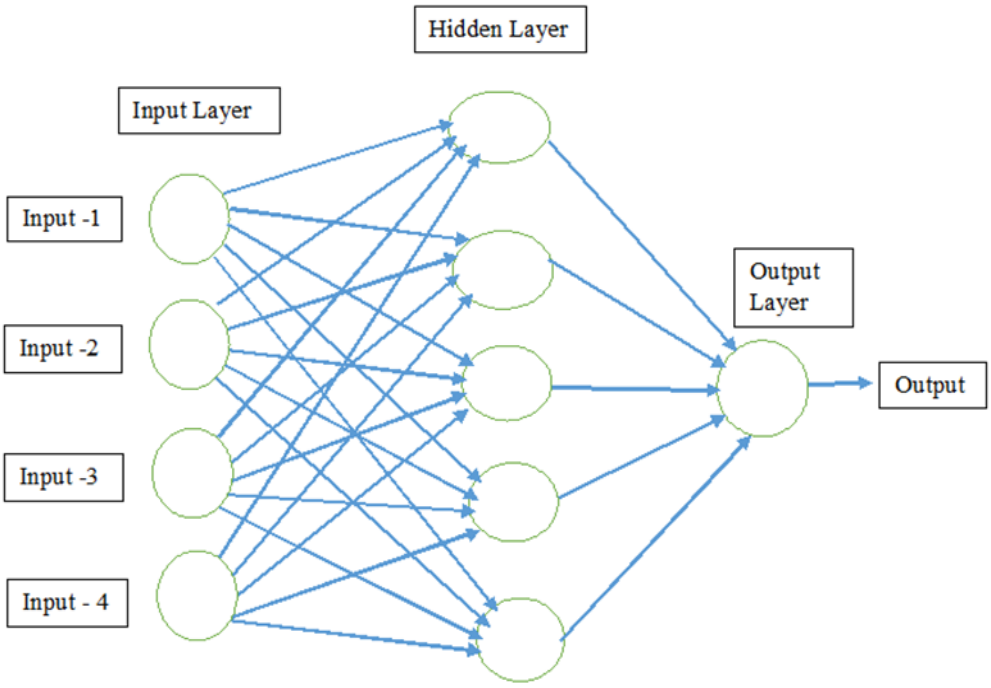
1. Предложен е нов метод за предварителна обработка на данните и премахване на шума от тях. Направено е сравнение със съвременни алгоритми по отношение на параметрите точност, прецизност и чувствителност. От резултатите е видно, че предложеният алгоритъм има по-добри характеристики.
2. Предложен е нов хибриден метод за подбор на характеристики. Методът е комбиниран с новопредложен алгоритъм за предварителна обработка на данните и модифициран Naive Bayes класификатор. Направените изследвания по отношение на точността на класификация, прецизността и чувствителността показват предимствата на метода в сравнение със съществуващите до момента методи.

# ГЛАВА 4. СИСТЕМА ЗА ДЕТЕКЦИЯ НА ИНТРУЗИИ ИЗПОЛЗВАЩА МОДИФИЦИРАН NAIVE BAYES КЛАСИФИКАТОР

Един от подходите за изграждане на системни модели е класификацията на входните данни постъпващи в системата. Оценката на поведението на алгоритмите е важна за избора на конкретен подход въз основа на поставените изисквания. В системите за детекция на интрузии базирани на използването на алгоритми за извличане на данни „data mining” методите за класификация са от решаващо значение. Основната цел на тази глава е да се предложат подходи за оценка на методите за класификация и да се потърсят начини за подобряване на тяхната работа.

## 4.1. Невронни мрежи

Изкуствените невронни мрежи (ANN) използват машинното обучение и когнитивната наука за да симулират поведението на невронните мрежи в природата. Те се използват за да се изчислят или апроксимират възможните решения, зависещи от голям набор от входни данни, които най-често са с голяма степен на неопределеност. Невронните мрежи представляват структури от свързани „неврони", които разменят съобщения помежду си. Например, невронна мрежа за разпознаване на написан на ръка текст се описва с набор от входни неврони, които се активират с помощта на пиксели от изображението. След като се претеглят и трансформират по определени характеристики, активирането на тези неврони се предава, директно на следващите след тях. Тези операции се повтарят, докато накрая се активира изходният неврон, който определя крайния резултат.




**Фигура 4.1. Структура на невронна мрежа**

На Фиг. 4.1 е показана структурата на невронна мрежа (НМ). Тя се състои в общия случай от входен, изходен и един или повече вътрешни слоеве. Всеки възел от входния слой е свързан към такъв от вътрешния, а те от своя страна с неврон от изходния. За всяка връзка има теглови коефициент.

**Брой възли и слоеве:** Изборът на различен брой на възлите за всеки слой се базира на сложността на задачата, която НМ се опитва да реши.

**Работа и обучение на НМ:** При своята работа мрежата отчита резултатите от предходния етап на обработка. Съществен етап от работата е периодът на обучение на мрежата. В рамките на този период резултатите на изхода на мрежата се сравняват с еталонни резултати за да може тя да се обучи.

**Алгоритъм „Back Propagation“ (BP):** Това е един от най-популярните алгоритми за реализация на НМ. Изследванията на този метод показват, че той може да бъде разделен на четири основни стъпки. След избирането, на случаен принцип, на теглови коефициенти на мрежата от определен набор се прилага алгоритъм с обратна връзка за да се изчислят необходимите корекции.

- **Приложения:** НМ намират приложение в най-различни сфери на живота. Във финансовата сфера се използват при класифицирането на кредитните рейтинги и прогнозите на пазара.
- **Недостатъци:** Изискват голям набор от еталонни входни данни за да се обучи ефективно мрежата.

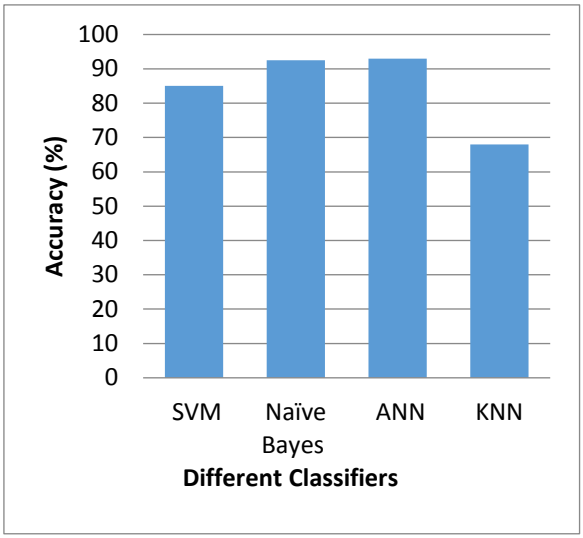
## 4.2 Naive Bayes класификатор

Naive Bayes е опростена техника за конструиране на класификатори, при която се присвояват етикети на отделните примери, представени като вектори от характерни стойности. Етикетите се извличат от определен брой крайни множества. Всички Naive Bayes класификатори приемат, че стойността на дадена характеристика е независима от стойността на другите. Техниката се основава на така наречената теорема на Bayes и е особено подходяща, когато размерността на входовете е висока. Въпреки своята простота, този класификатор често може да превъзхожда по-сложните методи за класификация.

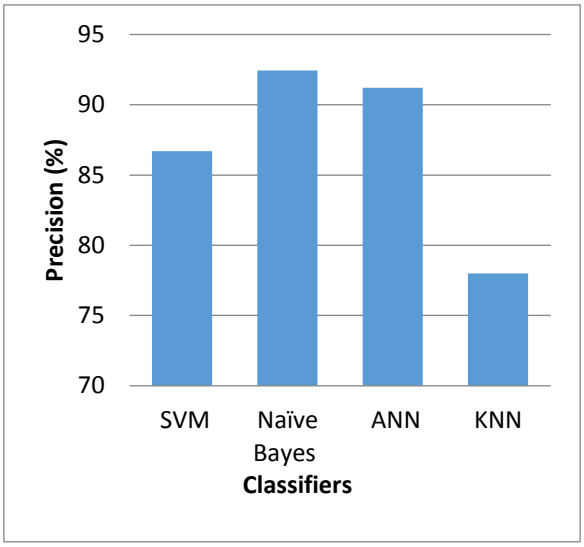
- **Преимущества:** Бързо обучение и ниска изчислителна сложност.
- **Недостатъци:** Приема се, че функциите са независими.

## 4.3. Резултати от анализа на различните класификатори

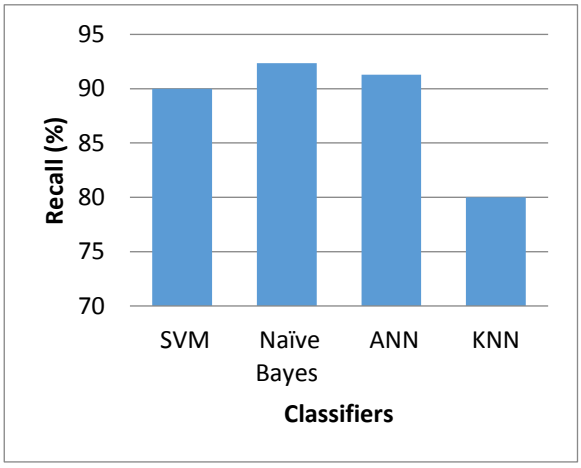
За да се оцени работата на различните класификатори е предложена и реализирана комбинация от новия алгоритъм за предварителна обработка, хибридния алгоритъм за избор на характеристики и ANN, KNN, SVM, Naive Bayes класификатори [A8]. Оценяват се техните възможности за откриване на атаки, като се използва трафик в реално време и данни от базата KDD CUP 99. Сравнението е по отношение на точност, прецизност и чувствителност.



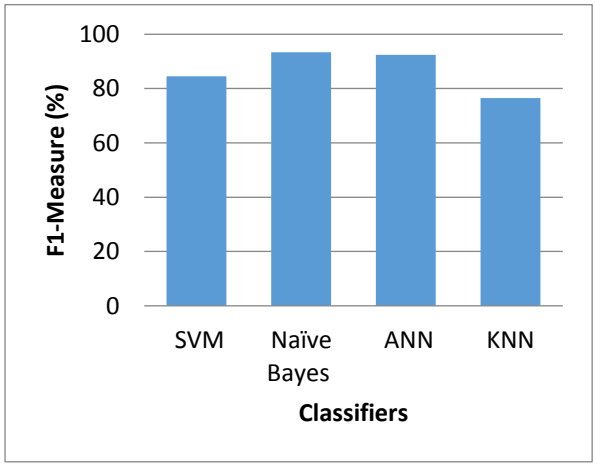

**Фигура 4.10. Оценка на точността**




**Фигура 4.11. Оценка на прецизността**




**Фигура 4.12. Оценка на чувствителността**




**Фигура 4.13. F1- оценка**

На Фиг. 4.10, 4.11, 4.12 и 4.13 са представени получените резултати. Фиг. 4.10 показва точността на откриване за всеки класификатор, използвайки предложените алгоритми за предварителна обработка и избор на характеристики. Вижда се че, работата на KNN класификатора не е задоволителна в сравнение с другите, докато ANN и Naive Bayes показват по-добра точност. Сравнявайки останалите параметри Naive Bayes класификаторът е най-добрият класификатор използван в комбинация с предложените алгоритми за предварителна обработка и избор на характеристики. Вторият и близък до резултатите на Naive Bayes класификатора е ANN класификаторът. Класификаторът KNN показва най-лоши резултати.

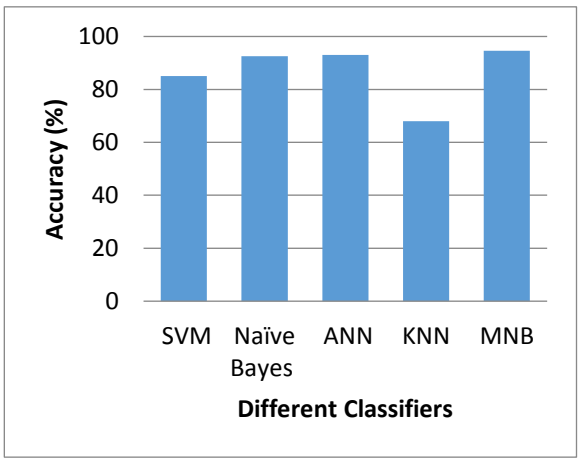
## 4.4. Модифициран Naive Bayes класификатор

Резултатите получени за различните класификатори са основание да се съсредоточат изследванията върху Naive Bayes класификатора. По този начин се дава възможност да се оптимизира работата на предложените алгоритми за предварителна обработка на данните и избор на характеристики. Горепосочените резултати доказват, че използването на подходящи класификатори оказва съществено влияние върху поведението на системата за детекция на интрузии. Тук Naive Bayes класификатора е модифициран за да се справи с разпознаването на различни атаки, като например U2R, R2L, DDoS, Prob [A9]. Чрез съвместното използване на нов алгоритъм за предварителна обработка на данни, алгоритъм за избор на характеристики и модифициран Naive Bayes класификатор (MNB) е разработена система за откриване на интрузии/атаки MNBIDS. По-долу са обобщени измененията на съществуващия Naive Bayes класификатор [A9]:

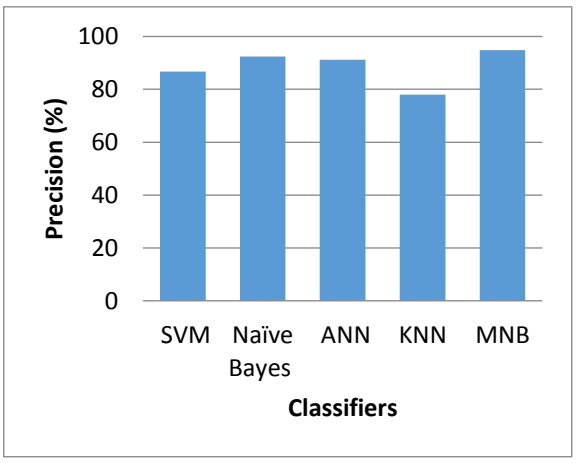
1. NB е модифициран, за да се намали неговата изчислителна сложност, така че да може ефективно да се прилага с прости оператори в процеса на откриване и класифициране.
2. MNB въвежда параметър за настройка между различните класове от обучаващи поредици, което позволява подобряване на точността на класификация.
3. Опростена е апроксимиращата експоненциална функция с линейна функция, което води до подобряване на производителността.

## 4.5. Анализ на поведението на MNBIDS

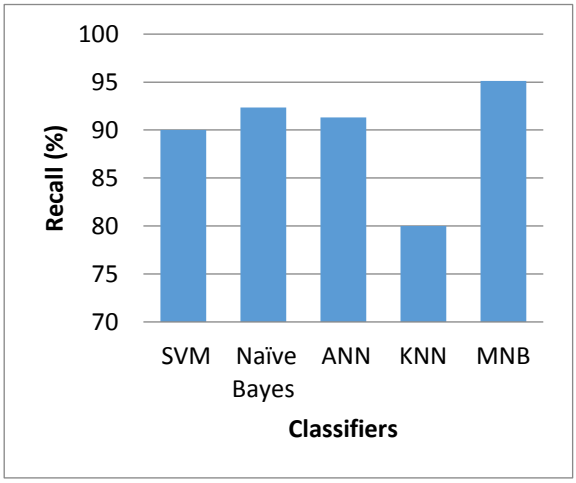
Направени са оценки на работата на предложената MNBIDS система с използване на различни класификатори, както и свързаните с тях методи [A9]. Резултатите доказват ефективността на предложената предварителна обработка и алгоритъма за извличане на характеристики когато се използва MNB класификатор.



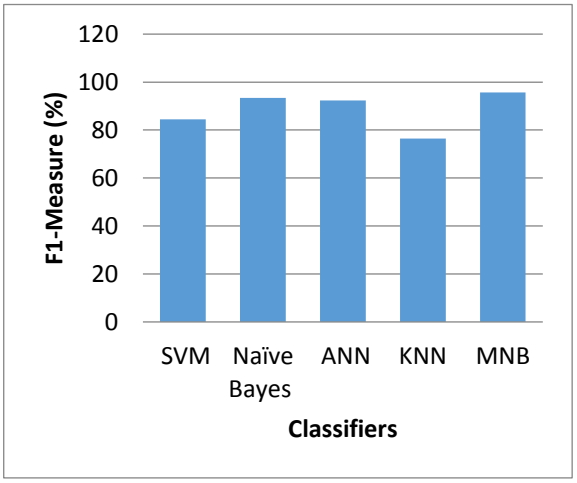

**Фигура 4.16. Оценка на точността**




**Фигура 4.17. Оценка на прецизността**




**Фигура 4.18. Оценка на чувствителността**




**Фигура 4.19. F1- оценка**

На Фиг. 4.16 е показана точността на откриване за всеки класификатор, използвайки предложените алгоритми за предварителна обработка и извличане на характеристики. MNB има най-добри резултати в сравнение с всички съществуващи класификатори. Аналогично на Фиг. 4.17, 4.18 и 4.19 се вижда, че MNB е показал най-добри резултати в сравнение с всички съществуващи класификатори по отношение на прецизност, чувствителност и F1-оценка.

Тъй като MNBIDS е комплексна система, която се състои от три основни компонента, а именно предварителна обработка на данни, избор на характеристики и модифициран MNB класификатор е важно да се сравни работата на системата MNBIDS с най-съвременните IDS системи. В този раздел са представени резултати от двата набора от данни, от базата на KDD CUP 99 и данни прихванати в реално време.

**Таблица 4.1. Оценка на прецизността**

| Метод | Стойност на параметъра прецизност (%) |
|---|---|
| W. Feng *et.al* [165] | 85.69 |
| Long-Sheng Chen *et.al* [166] | 89.44 |
| Yanjie Zhao *et.al* [167] | 86.48 |
| P. Lakshmi *et.al* [169] | 93.42 |
| **Proposed (MNBIDS)** | **96.77** |

**Таблица 4.2. Оценка на чувствителността**

| Метод | Стойност на параметъра чувствителност (%) |
|---|---|
| W. Feng *et.al* [165] | 88.72 |
| Long-Sheng Chen *et.al* [166] | 89.93 |
| Yanjie Zhao *et.al* [167] | 84.56 |
| P. Lakshmi *et.al* [169] | 92.34 |
| **Proposed (MNBIDS)** | **95.82** |

**Таблица 4.3. Оценка на точността**

| Метод | Стойност на параметъра точност (%) |
|---|---|
| W. Feng *et.al* [165] | 87.23 |
| Long-Sheng Chen *et.al* [166] | 88.92 |
| Yanjie Zhao *et.al* [167] | 87.93 |
| P. Lakshmi *et.al* [169] | 92.87 |
| **Proposed (MNBIDS)** | **96.35** |

Данните от таблици 4.1, 4.2 и 4.3 показват, че разработената нова система за откриване на интрузии/атаки превъзхожда значително съществуващите IDS модели поради използването на нов алгоритъм за предварителна обработка на данните, хибриден алгоритъм за избор на характеристики и модифициран NB класификатор.

## 4.6. ЗАКЛЮЧЕНИЕ

В тази част от дисертационния труд се извършват изследвания на различни класификатори като KNN, ANN, SVM и NB. Работата на всеки от тях е детайлно изучена и са показани техните предимства и недостатъци. След това, алгоритмите за предварителна обработка и избор на характеристики се прилагат върху масивите от данни в реално време и данни от базата KDDcup. Съгласно представените резултати е дефинирана целта за проектиране на нов класификатор с повишена ефективност. Това е реализирано чрез разработения модифициран Naive Bayes класификатор. Предложена е нова система за откриване на интрузии/атаки MNBIDS, която е сравнена по отношение на прецизност, чувствителност и точност със сходни системи IDS. Крайните резултати показват, че системата MNBIDS е постигнала подобрение по всички показатели в сравнение с другите използвани подходи.

### 4.7. Приноси към четвърта глава

1. Направена е оценка на ефективността на предложената предварителна обработка на наборите от данни и алгоритъма за избор на характеристики, използвайки различни класификатори като KNN, ANN, SVM и NB.
2. Модифициран е NB класификатора за да се разширят възможностите за откриване на атаки и класифициране с минимална сложност. С негова помощ е разработена система за откриване на атаки MNBIDS, която превъзхожда по основните показатели съществуващите IDS системи.

## III. ЗАКЛЮЧЕНИЕ И ОСНОВНИ ПРИНОСИ

Темата на този дисертационен труд е свързана със сигурните и ефективни подходи, модели и алгоритми за защита на информационни ресурси при едни от най-тежките атаки към компютърната сигурност – DDOS. Основната цел на дисертационния труд е разработването и формулирането на нови алгоритми и методи за предварителна обработка на данни, избор на характеристики и класификация на данни. С предложените алгоритми и методи да се проектира и оцени поведението на нова система за детекция на интрузии и да се подобрят съществуващите класификатори, за да се повиши ефективността на откриването на атаките. За да се изпълни целта са разработени различни техники и алгоритми, които осигуряват изпълнението на поставената цел. Анализът и проведените експерименти са на база сравнение и с вече съществуващи подходи за да се определи ефикасността на предложените решения.

Основните приноси на дисертационният труд са:

1. Анализиран е метод, базиран на TCP Congestion Window, за откриване на атаката Botnet. Анализът е направен, като е използван CUSUM алгоритъм реализиран върху платформа на мрежов симулатор. При този метод целта е да се оптимизират параметри, като най-кратко време за откриване, ниски стойности на вероятността за грешно откриване и малка изчислителна сложност.

2. Направен е анализ на метод за откриване на атаката Botnet чрез прилагане на машинно обучение. При него се използват техники базирани на Identifier / Locator подхода за откриване на DDOS атаки. Едно от най-големите преимущества на този метод е трудното осъществяване на контрол от страна на атакуващия върху мрежите Botnets.

3. Направен е анализ на характеристиките на появилите се напоследък DDOS атаки реализирани на приложния слой. Показано е влиянието, което оказват

този тип атаки. Изледвани са специфичните характеристики на DDOS атаките с наводняване на мрежата и методите за защита при този тип атаки.

4. Използвайки UML моделиране е разработен модел на DDOS атака реализирана на приложния слой. UML моделът ясно категоризира различните етапи на атаката. Целта е намаляване на времето за откриване на DDOS атаки на приложно ниво.

5. Разработен е алгоритъм, който дава възможност в реално време да се прихващат данни за DDOS атаки и едновременно с това да се намалява нивото на шума. При анализа на атаките може да има грешки при събирането на данни, и високо ниво на шум, дължащ се на неподходящо подбрани обекти от данни. Новият алгоритъм подобрява техниките за филтрирането на данни, като целта е да се намали времето за откриване.

6. Предложен е нов хибриден метод за подбор на характеристики (HFSA) при откриване на DDOS атаки. Методът е комбиниран с новопредложения алгоритъм за предварителна обработка на данните и модифициран Naive Bayes класификатор за да се разработи нова система за откриване на интрузии/атаки MNBIDS.

## IV. **ИЗПОЛЗВАНИ СЪКРАЩЕНИЯ**

| | |
|---|---|
| ARIMA | Autoregressive Integrated Moving Average |
| ANN | Artificial Neural Network |
| BP | Back Propagation |
| CN | Communication Networks |
| DNS | Domain Name System |
| DAAD | DNS Amplification Attacks Detector |
| DOS | Denial of Service |
| DDOS | Distributed Denial of Service |
| DARPA | Defense Advanced Research Projects Agency |
| FTP | File Transfer Protocol |
| HTTP | Hyper Text Transfer Protocol |
| HFSA | Hybrid Feature Selection Algorithm |
| HFSM | Hybrid Feature Selection Method () |
| IIDPS | Internal Intrusion Detection and Protection System |
| IP | Internet Protocol |
| IT | Information Technology |

ICMP Internet Control Message Protocol

KDD Knowledge Discovery & Data

KNN K-Nearest Neighbor

LAN Local Area Network

LOIC Low Orbit Ion Cannon

LSSVM-IDS Least Square Support Vector Machine –Intrusion Detection System

MNBIDS Modified Naïve Bayes Intrusion Detection System

NB Naïve Bayes

NIP Network Intrusion Prevention

OC-SVM One Class – Support Vector Machine

PCA Principal Component Analysis

R2L Remote to Local

RPC Remote Procedure Call

SVM Support Vector Machine

TCP/IP Transmission Control protocol/Internet Protocol

U2R User to Remote

UDP User Datagram Protocol

UML Unified Modeling Language

WAN Wireless Area Network

XML-DOS XML denial-of-service attack

## V. СПИСЪК НА АВТОРСКИТЕ ПУБЛИКАЦИИ

[A1] Karuna S. Bhosale, Maria Nenova and Georgi Iliev, "Review of Different Security Threats in Communication Networks", *International Journal of Academic Research and Development*, ISSN: 2455-4197, Volume 2, Issue 6, pp. 1253-1258, 2017.

[A2] Karuna S. Bhosale, Maria Nenova and Georgi Iliev, "Study and Analysis of Different Information and Resource Protection Methods in Communication Networks", *International Journal of Advanced Research and Development*, ISSN: 2455-4030, Volume 2, Issue 6, pp. 826-830, 2017.

[A3] Karuna S. Bhosale, Maria Nenova and Georgi Iliev, "Cyber Security Algorithms for Distributed Denial of Service Attack in Botnet", *in Proc. of Global Wireless Summit (GWS)*, Aarhus, Denmark, pp. 592-596, 27-30 Nov, 2016.

[A4] Karuna S. Bhosale, Maria Nenova and Georgi Iliev, "The Distributed Denial of Service Attacks (DDOS) Prevention Mechanisms on application layer**",** *2017 13th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS)*, Serbia, Nis, pp.136-139, October 18-20, 2017.

[A5] Karuna S. Bhosale, and Maria Nenova, "Detection of Application Layer DDOS Attacks Based on UML Modelling**",** *in International Conference on Smart City and Emerging Technologies*, Mumbai, India, pp.1-6, 5$^{Th}$ January, 2018.

[A6] Karuna S. Bhosale, Maria Nenova and Georgi Iliev, "Real Time Network Data Capturing using Attack Generators and Data Processing Method," in *International Conference on Computer Networks and Inventive Communications Technologies (ICCNCT-2018*), Part of the Lecture Notes on Data Engineering and Communications Technologies book series (LNDECT, volume 15), pp. 705-715, 26-27, Coimbatore, India, April 2018.

[A7] Karuna S. Bhosale, Maria Nenova and Georgi Iliev, "Data Mining Based Advanced Algorithm for Intrusion Detections in Communication Networks**",** *Presented in International Conference on Computational Techniques, Electronics & Mechanical Systems*, at Belgavi, India, 21-23 December, 2018.

[A8] Karuna S. Bhosale, Maria Nenova, and Georgi Iliev, "Intrusion Detection in Communication Networks Using Different Classifier", *Presented at 2$^{nd}$ International Conference On Advanced Technologies for Societal Applications*, at Pandharpur, India,14-15 December, 2018.

[A9] Karuna S. Bhosale, "Modified Naive Bayes Intrusion Detection System (MNBIDS)", Presented in *International Conference on Computational Techniques, Electronics & Mechanical Systems*, at Belgavi, India, 21-23 December, 2018**.**

## VI. ЛИТЕРУТУРНИ ИЗТОЧНИЦИ


[13] SANS Institute,"The Changing Face of Distributed Denial-of-Service Mitigation", 2001.

[14] Zhang Chao-yang, "DOS attack analysis and study of new measures to prevent," *International Conference on Intelligence Science and Information Engineering*, 2011

[18] Qijun Gu, and Peng Liu [online], "Denial of Service Attacks", 2012.

[19] Khaled M. Elleithy, Drazen Blagovic, Wang Cheng, and Paul Sideleau, "Denial of Service Attack Techniques: Analysis, Implementation and Comparison", *Systemics, Cybernetics and Informatics,* Volume 3 – Number -1, 2013.

[21] F Karami, Park and McCoy, "Stress Testing the Booters: Understanding and Undermining the Business of DDoS Services", August 2015.

[22] Adrien Bonguet and Martine Bellaiche, “A Survey of Denial of Service and Distributed Denial of Service And Defence in Cloud Computing”, *Future Internet*, vol. 9, no. 43, 2017.

[23] Khundrakpam Johnson Singh and Tanmay De, “An Approach of DDOS Attack Detection Using Classifiers”, *Communication Systems and Networks (COMSNETS),* 2011 Third International Conference on, 2011.

[24] E. T. Ferreira, G. A. Carrijo, R. de Oliveira and N. V. S. Araújo, “Intrusion Detection System with Wavelet and Neural Artifical Network Approach for Networks Computers”, *IEEE LATIN AMERICA TRANSACTIONS*, VOL. 9, NO. 5, SEPTEMBER 2011.

[26] Mohammed A. Ambusaidi, Xiangjian He, Priyadarsi Nanda and Zhiyuan Tan, “Building an intrusion detection system using a filter-based feature selection algorithm”, *IEEE TRANSACTIONS ON COMPUTERS,* VOL. NO 10, NOVEMBER 2014.

[27] Fang-Yie Leu, Kun-Lin Tsai, Me, Yi-Ting Hsiao, and Chao-Tung Yang, An Internal Intrusion Detection and Protection System by Using Data Mining and Forensic Techniques”, IEEE SYSTEMS JOURNAL, 2015.

[28] Seyyed Meysam Tabatabaie Nezhad, Mahboubeh Nazariy, and Ebrahim A. Gharavol, A Novel DoS and DDoS Attacks Detection Algorithm Using ARIMA Time Series Model and Chaotic System in Computer Networks," *IEEE Communications Letters*, Vol. 20, no. 4, 2016.

[33] Rajab M., A, J. Zarfoss, F. Monrose, and A. Terzis, “A multifaceted approach to understanding the botnet phenomenon,” in *Proceedings of the 6th Internet Measurement Conference*, 2006.

[34] Cooke E, F. Jahanian, and D. McPherson, “The zombie roundup: Understanding, detecting, and disrupting botnets,” in *Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI‘05)*, pp. 39–44. USENIX, June 2005.

[37] Baecher P, Koetter, M, Holz, T, Dornseif M, & Freiling F, “The nepenthes platform: An efficient approach to collect malware,” in *International Workshop on Recent Advances in Intrusion Detection* , *Springer*, pp. 165–184, 2006.

[38] Holz T, Steiner M, Dahl F, Biersack E, & Freiling F. C, “Measurements and Mitigation of Peer-to-Peer based Botnets: A Case Study on Storm Worm,” *LEET*, 8 (1), 1-9, 2008.

[127] Indraneel Sreeram, and Venkata Praveen KumarVuppala, “HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm”, *Applied Computing and Informatics*, Volume 15, Issue 1, pp. 59-66, January 2019.

[128] Esraa Alomari, 2Selvakumar Manickam, 3,4B. B. Gupta, 5Shankar Karuppayah, 6Rafeef Alfaris, "Botnet-based Distributed Denial of Service (DDOS) Attacks on Web Servers: Classification and Art**"** *International Journal of Computer Applications pp. 0975 – 8887, Volume 49– No.7, July 2012.*

[139] Chuibi Huang1, Jinlin Wang1, 2, Gang Wu1, Jun Chen2, "Mining Web User Behaviors to Detect Application Layer DDOS Attacks", *Journal of Software*, vol. 9, no. 4, April 2014.

[140] *Karanpreet Singh a, Paramvir Singh a, Krishan Kumar b,* "Application layer HTTP-GET flood DDOS attacks: Research landscape and challenges", *Elsevie*r, 2016.

[156] H. Xiong; Gaurav Pandey; M. Steinbach; Vipin Kumar, "Enhancing data analysis with noise removal", IEEE Transactions on Knowledge and Data Engineering, Volume: 18, Issue: 3, March 2006.

[158] Muhammad Fahim Uddin, "Noise Removal and Structured Data Detection to improve search for personality features", Advances in Social Networks Analysis and Mining (ASONAM), 2016.

[164] K. Uysal, and S. Gunal, "A novel probabilistic feature selection method for text classification," Knowledge-Based Systems, 36, pp. 226-23, 2012.

[165] W. Feng, Q. Zhang, G. Uu, J. X. Huang, "Mining network data for intrusion detection through combining SVMs with ant colony networks", Future Generation Computer Systems 37, 2014.

[166] Long-Sheng Chen, Jhih-Siang Syu, "Feature Extraction based Approaches for Improving the Performance of Intrusion Detection Systems". Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol I, IMECS Hong Kong, March 18 - 20, 2015.

[167] Yanjie Zhao, "Network intrusion detection system model based on data mining", Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2016.

[169] P. Lakshmi, and D. Geetha, "Intrusion Detection System Using Modified Support Vector Machine", Networking and Communication Engineering, vol.7, no.10, 2015.

# TECHNICAL UNIVERSITY OF SOFIA

## FACULTY OF TELECOMMUNICATIONS

DEPARTMENT OF "COMMUNICATION NETWORKS"

**Karuna Suryakant Bhosale**


## Investigations of Algorithms and Methods for Information and Resource Protection in Communication Networks

## ABSTRACT of Ph.D THESIS


The topic of this thesis is related to investigations of algorithms and methods for information and resource protection in communication networks, with the goal defined as the "development and formulation of novel algorithms and methods for data pre-processing, features selection/construction and classification to enhance the security performance in communication networks, also to design and evaluate a new IDS framework and to modify the existing classifiers in order to enhance the classification performance".

The thesis contains 4 chapters, 52 figures and 5 tables. 169 references are used with 75% of them from last 10 years. The analysis, proposed methods and results are presented through 9 author publications, 2 of which are in international scientific journals and 7 are presented at international conferences. Chapter 1 provides a survey of information and resource protection methods for communication networks. Chapter 2 emphasizes on the analysis of characteristics of newly designed application layer distributed denial of service attacks. A new model using UML modeling for application layer DDOS attacks is proposed. Chapter 3 proposes the data cleaning method and evaluation of performance against state-of-art solutions is presented. Also, a novel features extraction method is proposed called Hybrid Feature Selection Algorithm (HFSA). Chapter 4 proposes the performance evaluation of proposed pre-processing and feature selection algorithms using the different classifiers such as KNN, ANN, SVM, and NB. Modified NB classifier is developed to enhance the attack detection and classification performance with minimum complexity.

The major contributions of the thesis are listed below. 1) Analysis of characteristics of newly designed application layer distributed denial of service attacks is performed. 2) With the help of UML modeling, a model for application layer DDOS attack is proposed. 3) Real-time data capturing with the DDOS attack along with data preprocessing noise removal algorithm is proposed. 4) A novel features extraction method is proposed called Hybrid Feature Selection Algorithm (HFSA). HFSA is then combined with the proposed pre-processing and the Naive Bayes classifier. The classification performance is measured in terms of classification accuracy, precision, recall and compared with recent intrusion detection methods.