

Instituto Politécnico de Leiria  
Escola Superior de Tecnologia e Gestão  
Departamento de Engenharia Informática  
Mestrado em Cibersegurança e Informática Forense

INTELIGÊNCIA ARTIFICIAL APLICADA À  
DETEÇÃO DE INCIDENTES DE SEGURANÇA EM  
REDES IOT

TIÉZER COSTA DE MELO

Leiria, Março de 2023

PREVIEW

Instituto Politécnico de Leiria  
Escola Superior de Tecnologia e Gestão  
Departamento de Engenharia Informática  
Mestrado em Cibersegurança e Informática Forense

INTELIGÊNCIA ARTIFICIAL APLICADA À  
DETEÇÃO DE INCIDENTES DE SEGURANÇA EM  
REDES IOT

TIÉZER COSTA DE MELO

Número: 2200175

Dissertação realizada sob orientação do Professor Doutor Carlos Manuel da Silva Rabadão ([carlos.rabadao@ipleiria.pt](mailto:carlos.rabadao@ipleiria.pt)), do Professor Doutor Leonel Filipe Simões Santos ([leonel.santos@ipleiria.pt](mailto:leonel.santos@ipleiria.pt)) e do Professor Doutor Rogério Luís de Carvalho Costa ([rogerio.l.costa@ipleiria.pt](mailto:rogerio.l.costa@ipleiria.pt)).

Leiria, Março de 2023

PREVIEW

## AGRADECIMENTOS

---

Primeiramente, quero gostaria de agradecer ao meu Deus, que me sustentou em todos os momentos de dificuldade e sempre se manteve fiel.

Gostaria de agradecer também aos meus pais, Cláudia e Gilberto, que são os melhores pais do mundo, que não mediram esforços para me proporcionar a realização de um sonho e sempre que necessário, traziam palavras de conforto. Também, a minha avó, que todos os dias demonstra o que é o amor em forma de pessoa. Agradecer também a Kelen, que mesmo com a distância, sempre se manteve presente, trazendo palavras de incentivo e algumas verdades, quando necessário.

Agradeço aos meus orientadores, Carlos, Leonel e Rogério, pela extrema paciência e solicitude em meio a inúmeros questionamentos e curiosidades e por todo os ensinamentos ao longo deste trabalho.

Aos meus colegas, Rafaela e António, e ao meu professor Professor Doutor Miguel Frade, agradeço pela oportunidade de ter participado de inúmeros momentos inusitados em meio às aulas e trabalhos académicos. Aos meus colegas Rafael Ascensão e João Cardo pelo coleguismo nas atividades profissionais.

Por fim, mas não menos importante, agradeço também aos meus amigos e colegas por, de alguma forma, terem participado desta pequena e desafiadora fase da minha vida e terem contribuído para a conclusão dela.

PREVIEW

## RESUMO

---

*Internet of Thing* ou IoT são dispositivos de limitado poder computacional, interconectados através da internet ou outra rede de comunicação, que partilham informação entre si e atuam de forma autónoma com uma mínima intervenção humana. Devido a algumas destas características, eles têm sido utilizados em diversas áreas da sociedade. Porém, apesar dos diversos benefícios trazidos por este tipo de dispositivos, estes apresentam alguns problemas de segurança. Tais problemas surgem devido à sua menor capacidade computacional, que impede a aplicação de técnicas de proteção mais complexas, e à grande diversidade ou heterogeneidade de tecnologias utilizadas (*hardware*, protocolos etc.). Como alternativa, técnicas de *machine learning* (ML) tem sido aplicadas como forma de melhorar a capacidade de deteção de ataques e tráfego anómalo. Neste trabalho foram criados dois *datasets* com intuito de representar os serviços de uma *smart greenhouse* e um conjunto de apartamentos que utilizam sistemas inteligentes de controlo. Os *datasets*, que representam o tráfego de dados destas duas redes IoT, são compostos pelos protocolos CoAP e MQTT. Foi realizada a revisão e a análise das ferramentas de simulação e geração de tráfego IoT, onde, através de comparação das características, foram selecionadas as ferramentas Contiki e Netsim. As simulações foram executadas através destas duas ferramentas e, além do tráfego normal, foram simulados 6 diferentes ataques, cuja maior parte destes estava relacionado com o protocolo RPL. Aos dados destes *datasets*, foram aplicados modelos de aprendizagem de máquina com o intuito de identificar os ataques utilizados, onde foi obtido um alto índice de acerto no que se refere à classificação do tráfego malicioso.

PREVIEW



## ABSTRACT

---

Internet of Thing or IoT are devices with limited computational power, interconnected via the internet or another communication network, which share information with each other and act autonomously with minimal human intervention. Due to some of these characteristics, they have been used in different areas of society. However, despite the many benefits brought by this type of devices, they have some security problems. Such problems arise due to its lower computational capacity, which prevents the application of more complex protection techniques, and to the great diversity or heterogeneity of technologies used (hardware, protocols, etc.). As an alternative, machine learning (ML) techniques have been applied as a way to improve the ability to detect attacks and anomalous traffic. In this work, two datasets were created in order to represent the services of a smart greenhouse and a set of apartments that use intelligent control systems. The datasets, which represent the data traffic of these two IoT networks, are composed of the CoAP and MQTT protocols. A review and analysis of the IoT simulation and traffic generation tools was carried out, where, by comparing the characteristics, the Contiki and Netsim tools were selected. The simulations were performed using these two tools and, in addition to normal traffic, 6 different attacks were simulated, most of which were related to the RPL protocol. Machine learning models were applied to the data from these datasets in order to identify the attacks used, where a high success rate was obtained regarding the classification of malicious traffic.

PREVIEW

## ÍNDICE

---

Agradecimentos	i
Resumo	iii
Abstract	v
Índice	vii
Lista de Figuras	xi
Lista de Tabelas	xiii
Lista de Abreviaturas	xvii
1 INTRODUÇÃO	1
1.1 Objetivos e contribuições	3
1.2 Estrutura do trabalho	3
2 BACKGROUND	5
2.1 Internet of Things	5
2.1.1 Caracterização de Internet of Things	5
2.1.2 Componentes	6
2.1.3 Arquiteturas	7
2.1.4 Protocolos	9
2.1.5 Segurança, ameaças e vulnerabilidades	11
2.1.6 Ataques	12
2.1.7 Smart Farming	14
2.1.8 Smart Cities	17
2.2 Inteligencia Artificial	18
2.3 Intrusion Detection System	22
2.4 Datasets	24
2.5 Trabalhos Relacionados	27
3 SOLUÇÕES DE GERAÇÃO DE DADOS	31

3.1	Soluções de Geração de Dados . . . . .	31
3.2	Características de interesse . . . . .	31
3.3	Contiki/Cooja . . . . .	33
3.3.1	Características e Funcionalidades . . . . .	33
3.3.2	Avaliação das Características . . . . .	34
3.4	Network Simulator 3 . . . . .	34
3.4.1	Características e Funcionalidades . . . . .	34
3.4.2	Avaliação das Características . . . . .	35
3.5	IoT <i>Dataset</i> Generator Framework . . . . .	35
3.5.1	Características e Funcionalidades . . . . .	35
3.5.2	Avaliação das Características . . . . .	36
3.6	IoT-Flock . . . . .	37
3.6.1	Características e Funcionalidades . . . . .	37
3.6.2	Avaliação das Características . . . . .	37
3.7	MQTT Generator . . . . .	38
3.7.1	Características e Funcionalidades . . . . .	38
3.7.2	Avaliação das Características . . . . .	38
3.8	COAP Protocol Simulator . . . . .	39
3.8.1	Características e Funcionalidades . . . . .	39
3.8.2	Avaliação das Características . . . . .	39
3.9	Scapy . . . . .	39
3.9.1	Características e Funcionalidades . . . . .	40
3.9.2	Avaliação das Características . . . . .	40
3.10	Netsim . . . . .	41
3.10.1	Características e Funcionalidades . . . . .	41
3.10.2	Avaliação das Características . . . . .	42
3.11	Resumo Comparativo das Ferramentas . . . . .	42
3.12	Definição da Ferramenta . . . . .	42
3.13	Soluções de geração de fluxo de dados . . . . .	44
3.14	Análise das ferramentas de geração de fluxo . . . . .	44
4	DEFINIÇÃO E CARACTERIZAÇÃO DO AMBIENTE E SIMULAÇÃO . . . . .	47
4.1	Características e definições gerais . . . . .	47
4.2	Definição do âmbito das simulações . . . . .	48
4.3	Características tecnológicas das simulações . . . . .	50
4.4	Caracterização dos Ambientes Propostos . . . . .	51
4.4.1	Composição da rede na ferramenta Netsim . . . . .	52
4.4.2	Composição da rede na ferramenta Contiki . . . . .	53

4.5	Características da Simulação . . . . .	53
4.6	Temporização das simulações . . . . .	54
4.7	Ataques Seleccionados . . . . .	56
4.8	Limitações das ferramentas . . . . .	58
4.9	Versões dos Softwares utilizados . . . . .	60
5	SIMULAÇÕES E DATASETS . . . . .	61
5.1	Alterações realizadas nas ferramentas de simulação . . . . .	61
5.1.1	Alterações necessárias na ferramenta Contiki . . . . .	62
5.1.2	Alterações necessárias na ferramenta Netsim . . . . .	62
5.2	Preparação do ambiente da ferramenta Contiki . . . . .	63
5.2.1	Configuração da simulação na ferramenta Contiki . . . . .	65
5.2.2	Configuração da simulação na ferramenta Netsim . . . . .	66
5.3	Realização das simulações . . . . .	67
5.4	Captura do tráfego e geração dos ficheiros PCAP . . . . .	67
5.5	Definição dos filtros de protocolos . . . . .	68
5.6	Pré-processamento de dados . . . . .	69
5.6.1	Conversão de ficheiros PCAP em CSV . . . . .	70
5.6.2	Separação dos dados . . . . .	71
5.6.3	Limpeza dos dados . . . . .	72
5.6.4	Adição dos atributos-alvo . . . . .	72
5.6.5	Eliminação seletiva de atributos . . . . .	74
5.6.6	Geração dos <i>datasets</i> finais . . . . .	74
5.7	Descrição dos <i>datasets</i> gerados . . . . .	75
5.7.1	Descrição do <i>dataset smart_greenhouse</i> . . . . .	75
5.7.2	Descrição do <i>dataset smart_city</i> . . . . .	79
6	APLICAÇÃO DE ALGORITMOS DE MACHINE LEARNING . . . . .	83
6.1	Seleção dos algoritmos . . . . .	86
6.2	Técnicas de predição . . . . .	86
6.3	Eliminação de dados . . . . .	87
6.4	Separação dos dados . . . . .	88
6.5	Tratamento dos dados . . . . .	88
6.6	Normalização dos dados . . . . .	89
6.7	Codificação de variáveis categóricas . . . . .	90
6.8	Redução do domínio de dados . . . . .	91
6.9	Identificação dos melhores hiperparâmetros . . . . .	92
6.10	Caracterização dos algoritmos de classificação . . . . .	93

6.11 Métricas de Avaliação . . . . .	94
6.12 Classificação Binária . . . . .	97
6.12.1 smart_city-binary . . . . .	97
6.12.2 smart_greenhouse-binary . . . . .	100
6.13 Classificação Multi-classe . . . . .	103
6.13.1 smart_city-multiclass . . . . .	104
6.13.2 smart_greenhouse-multiclass . . . . .	108
6.14 Resultados . . . . .	114
 7 CONCLUSÕES . . . . .	 119
7.1 Contributos . . . . .	121
7.2 Trabalhos Futuros . . . . .	122
 BIBLIOGRAFIA . . . . .	 123
 DECLARAÇÃO . . . . .	 127

## LISTA DE FIGURAS

---

Figura 1	Representação das arquiteturas IoT baseadas em 3, 4, 5 e 6 camadas. . . . .	10
Figura 2	Representação das tecnologias e disposição dos sensores no ambiente do conjunto de apartamentos. . . . .	48
Figura 3	Representação das tecnologias e disposição dos sensores no ambiente da <i>smart greenhouse</i> . . . . .	49
Figura 4	Configuração no componente para geração do ficheiro PCAP na ferramenta Netsim. . . . .	52
Figura 5	Topologias de rede das simulações com 3 sensores realizadas no Netsim. . . . .	54
Figura 6	Topologias de rede das simulações com 9 sensores realizadas no Netsim. . . . .	54
Figura 7	Topologias de Rede das simulações realizadas no Contiki. . .	55
Figura 8	Ambiente da simulação do ataque DoS na ferramenta Netsim.	57
Figura 9	Ambiente da simulação do ataque <i>Sinkhole</i> na ferramenta Netsim. . . . .	57
Figura 10	Ambiente da simulação do ataque DIO Supression na ferramenta Netsim. . . . .	57
Figura 11	Erro apresentado na simulação do ataque <i>RPL DIS Flooding</i> quando habilitada a opção de gerar os ficheiros PCAP. . . .	58
Figura 12	Topologias de rede das simulações dos ataques realizados com o protocolo MQTT no Contiki. . . . .	59
Figura 13	Topologias de rede das simulações dos ataques realizados com o protocolo CoAP no Contiki. . . . .	59
Figura 14	Erro apresentado ao tentar inicializar uma simulação em um <i>branch</i> diferente ao do ataque selecionado. . . . .	64
Figura 15	Lista de atributos e respetivos tipos de dados do <i>dataset smart_greenhouse</i> . . . . .	78
Figura 16	Lista de atributos e respetivos tipos de dados do <i>dataset smart_city</i> . . . . .	81
Figura 17	Exemplo do <i>output</i> da função <i>classification_report</i> . . . . .	96

Figura 18	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_city-binary</i> com o conjunto de atributos número 1. . . . .	99
Figura 19	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_city-binary</i> com o conjunto de atributos número 2. . . . .	100
Figura 20	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_greenhouse-binary</i> com o conjunto de atributos número 1. . . . .	103
Figura 21	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_greenhouse-binary</i> com o conjunto de atributos número 3. . . . .	104
Figura 22	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_greenhouse-binary</i> com o conjunto de atributos número 5. . . . .	105
Figura 23	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_city-multiclass</i> com o conjunto de atributos número 1. . . . .	107
Figura 24	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_city-multiclass</i> com o conjunto de atributos número 2. . . . .	108
Figura 25	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_city-multiclass</i> com o conjunto de atributos número 5. . . . .	110
Figura 26	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_greenhouse-multiclass</i> com o conjunto de atributos número 1. . . . .	112
Figura 27	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_greenhouse-multiclass</i> com o conjunto de atributos número 2. . . . .	113
Figura 28	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_greenhouse-multiclass</i> com o conjunto de atributos número 3. . . . .	114
Figura 29	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_greenhouse-multiclass</i> com o conjunto de atributos número 4. . . . .	115
Figura 30	Matriz de confusão dos algoritmos aplicados ao <i>dataset smart_greenhouse-multiclass</i> com o conjunto de atributos número 5. . . . .	116



## LISTA DE TABELAS

---

Tabela 1	Comparação das características das ferramentas de geração de dados IoT . . . . .	42
Tabela 2	Lista dos <i>softwares</i> utilizados na simulação e respectivas versões. . . . .	60
Tabela 3	Valor respetivo a cada classe nos atributos <i>IS_MALICIOUS</i> e <i>ATTACK_TYPE</i> . . . . .	73
Tabela 4	Número de ficheiros para cada tipo de simulação realizada na ferramenta Contiki . . . . .	76
Tabela 5	Endereços IP maliciosos para cada tipo de ataque nas simulações realizadas na ferramenta Contiki . . . . .	77
Tabela 6	Número de registos normais e maliciosos do <i>dataset smart_greenhouse</i> . . . . .	78
Tabela 7	Número de ficheiros para cada tipo de simulação realizada na ferramenta Netsim. . . . .	79
Tabela 8	Endereços IP maliciosos para cada tipo de ataque nas simulações realizadas na ferramenta Netsim. . . . .	79
Tabela 9	Número de registos normais e maliciosos do <i>dataset smart_city</i> . . . . .	80
Tabela 10	Números de registos dos dois novos <i>datasets</i> criados a partir do <i>dataset smart_greenhouse</i> com base no tipo de classificação. . . . .	84
Tabela 11	Números de registos por tipo de tráfego nos <i>datasets</i> baseados no <i>smart_greenhouse</i> com base no tipo de classificação. . . . .	85
Tabela 12	Números de registos dos dois novos <i>datasets</i> criados a partir do <i>dataset smart_greenhouse</i> com base no tipo de classificação. . . . .	85
Tabela 13	Números de registos por tipo de tráfego nos <i>datasets</i> baseados no <i>smart_city</i> com base no tipo de classificação. . . . .	85
Tabela 14	Distribuição dos dados de treino e teste em cada <i>dataset</i> . . . . .	88
Tabela 15	Sensibilidade dos algoritmos quanto a variância na escala dos dados. . . . .	89
Tabela 16	Número identificador do conjunto de atributos e o algoritmo utilizado na seleção destes atributos. . . . .	92
Tabela 17	Número de registos de cada classe no conjunto de teste do <i>dataset smart_city-binary</i> quando usada a classificação binária. . . . .	97

Tabela 18	Conjunto de atributos selecionados pelos algoritmos de seleção de atributos a serem usados no <i>dataset smart_city-multiclass</i> . . . . .	98
Tabela 19	Resultados das métricas em relação ao <i>dataset smart_city-binary</i> . . . . .	98
Tabela 20	Resultados do desempenho dos algoritmos aplicados ao <i>dataset smart_city-binary</i> de acordo com as métricas de avaliação selecionadas. . . . .	100
Tabela 21	Número de registros de cada classe no conjunto de teste do <i>dataset smart_greenhouse-binary</i> quando usada a classificação binária. . . . .	100
Tabela 22	Conjunto de atributos selecionados pelos algoritmos de seleção de atributos a serem usados no <i>dataset smart_city-binary</i> .101	
Tabela 23	Resultados das métricas em relação ao <i>dataset smart_smartgreenhouse-binary</i> . . . . .	102
Tabela 24	Resultados do desempenho dos algoritmos aplicados ao <i>dataset smart_greenhouse-binary</i> de acordo com as métricas de avaliação selecionadas. . . . .	103
Tabela 25	Número de registros de cada classe no conjunto de teste do <i>dataset smart_city-multiclass</i> quando usada a classificação multi-classe. . . . .	105
Tabela 26	Conjunto de atributos selecionados pelos algoritmos de seleção de atributos a serem usados no <i>dataset smart_city-multiclass</i> . . . . .	106
Tabela 27	Resultados das métricas em relação ao <i>dataset smart_city-multiclass</i> . . . . .	106
Tabela 28	Resultados do desempenho dos algoritmos aplicados ao <i>dataset smart_city-multiclass</i> de acordo com as métricas de avaliação selecionadas. . . . .	108
Tabela 29	Número de registros de cada classe no conjunto de teste do <i>dataset smart_grenhouse-multiclass</i> quando usada a classificação multi-classe. . . . .	109
Tabela 30	Conjunto de atributos selecionados pelos algoritmos de seleção de atributos a serem usados no <i>dataset smart_greenhouse-multiclass</i> . . . . .	109
Tabela 31	Resultados das métricas em relação ao <i>dataset smart_greenhouse-multiclass</i> . . . . .	111

Tabela 32	Resultados do desempenho dos algoritmos aplicados ao <i>dataset smart_smartgreenhouse-multiclass</i> de acordo com as métricas de avaliação selecionadas. . . . .	113
-----------	---	-----

PREVIEW

PREVIEW

## LISTA DE ABREVIATURAS

---

6LowPAN IPv6 over Low-Power Wireless Personal Area Networks.

AIDS Anomaly-based Intrusion detection systems.

AMQP Advanced Message Queuing Protocol.

API Application Programming Interface.

APT advanced persistent threat.

AWS Amazon Web Services.

BLE Bluetooth Low Energy.

BSD Berkeley Source Distribution.

CIDS Colaborative Intrusion detection systems.

CO<sub>2</sub> Dióxido de Carbono.

CoAP Constrained Application Protocol.

CPU Central Processing Unit.

CSV Comma-separated values.

DDoS Distributed Denial of Service.

DL Deep Learning.

DNS Domain Name System.

DoS Denial of Service.

DTC Decision Tree Classifier.

FN false negative.

FP false positive.

GA	Genetic Algorithm.
GPRS	General Packet Radio Service.
GPS	Global Positioning System.
GSM	Global System for Mobile communication.
GUI	graphical user interface.
HIDS	Host-based intrusion detection system.
HTTP	Hypertext Transfer Protocol.
HTTPS	Hypertext Transfer Protocol Secure.
IA	Inteligência Artificial.
IBM	International Business Machines Corporation.
IDGF	IoT Dataset Generation Framework.
IDS	Intrusion detection systems.
IETF	The Internet Engineering Task Force.
IIoT	Industrial internet of things.
IOT	Internet of Things.
IP	Internet Protocol.
IPFIX	Internet Protocol Flow Information Export.
IPv4	Internet Protocol versão 4.
IPv6	Internet Protocol versão 6.
JSON	JavaScript Object Notation.
KNN	K-neighbors Classifier.
LAN	Local Area Networks.
LPWAN	Low-power, wide-area network.
LP-WAN	Low-power, wide-area network.

LRC	Logistic Regression Classifier.
LTE	Long Term Evolution.
MITM	Man-in-the-middle attack.
ML	Machine Learning.
MQTT	Message Queuing Telemetry Transport.
MQTT-SN	MQTT for Sensor Networks.
NAT	Network address translation.
NFC	Near field communication.
NIDS	Network-based intrusion detection system.
NoSQL	Non SQL ou not only SQL.
OSI	Open Systems Interconnection.
PCAP	Packet Capture.
pH	Potential of hydrogen.
PKI	Public key infrastructure.
QUIC	Quick UDP Internet Connections.
RFC	Random Forest Classifier.
RFID	Radio-frequency identification.
RNA	Redes Neurais Artificiais.
RPL	IPv6 Routing Protocol for Low-Power and Lossy Networks.
RSSF	Redes de Sensores sem-fio.
SIDS	Signature-based Intrusion detection systems.
SO	Sistema Operativo.

Lista de Abreviaturas

TCP	Transmission Control Protocol.
TIC	Tecnologia da Informação e Comunicação.
TLS	Transport Layer Security.
TN	true negative.
TP	true positive.
UART	universal asynchronous receiver / transmitter.
UDP	User Datagram Protocol.
UWB	Ultra-wideband.
WAN	Wide Area Networks.
WSN	wireless sensor networks.
XML	extensible markup language.
XMPP	Extensible Messaging and Presence Protocol.
YAF	Yet Another Flowmeter.