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

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), do Professor Doutor Leonel Filipe Simões Santos (leonel.santos@ipleiria.pt) e do Professor Doutor Rogério Luís de Carvalho Costa (rogerio.l.costa@ipleiria.pt).

Leiria, Março de 2023

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.

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 ferrametas 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.

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.

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABELAS

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.
CO2	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 Neuronais 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.

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 Netwoks.
WSN	wireless sensor networks.
XML	extensible markup language.
XMPP	Extensible Messaging and Presence Protocol.
YAF	Yet Another Flowmeter.