Utilizzo di ndpiReader e analisi statistica per il rilevamento di anomalie di una rete ospedaliera

Angelo Quartarone[1], Tommaso Tocchini[2], and Gabriele Scannagatti[3]

[1]a.quartarone1@studenti.unipi.it
[2]t.tocchini@studenti.unipi.it
[3]g.scannagatti@studenti.unipi.it
[1,2,3]Dipartimento di informatica, Università di Pisa

20 settembre 2024

Sommario

Questo progetto si concentra sull'analisi statistica del traffico di rete all'interno di un'infrastruttura ospedaliera utilizzando ndpiReader. L'obiettivo principale è identificare potenziali attacchi informatici, con particolare attenzione al protocollo MQTT, spesso utilizzato nei dispositivi IoT (Internet of Things) in ambito sanitario. A tal fine, sono state sviluppate feature che sfruttano l'Interquartile Range (IQR) come misura statistica per la rilevazione di anomalie nel traffico di rete, al fine di identificare outliers che potrebbero indicare attività malevole.

1 Introduzione

Negli ultimi anni, l'integrazione di dispositivi IoT (Internet of Things) nelle infrastrutture ospedaliere ha portato significativi miglioramenti nella gestione e nell'erogazione dei servizi sanitari. Dispositivi come sensori, monitor per i pazienti e sistemi di automazione migliorano l'efficienza e la qualità delle cure, ma allo stesso tempo introducono nuove vulnerabilità. Il protocollo MQTT, ampiamente utilizzato per la comunicazione tra dispositivi IoT, è particolarmente suscettibile ad attacchi informatici a causa della sua natura leggera e della mancanza di robuste misure di sicurezza integrate.

In un contesto critico come quello sanitario, la protezione delle reti è fondamentale per garantire la continuità operativa, la riservatezza dei dati e la sicurezza dei pazienti. Tuttavia, la complessità e la dinamicità del traffico di rete ospedaliero richiedono soluzioni avanzate per rilevare tempestivamente le minacce, soprattutto quelle che utilizzano protocolli specifici come MQTT.

L'obiettivo di questo progetto è implementare un sistema di analisi e rilevamento delle anomalie nel traffico di rete utilizzando ndpiReader; uno strumento potente per l'ispezione del traffico di rete, in grado di identificare e classificare un'ampia gamma di protocolli, tra cui MQTT. Per aumentare l'efficacia del sistema di rilevamento, è stata sviluppata un'analisi statistica basata sull'Interquartile Range (IQR), un metodo comunemente utilizzato per individuare outliers, ovvero dati anomali che potrebbero essere indicativi di attività malevola.

1.1 MQTT

MQTT (Message Queuing Telemetry Transport) è un protocollo di messaggistica PUB/SUB basato su TCP/IP, progettato per facilitare la comunicazione tra dispositivi con risorse limitate e reti a bassa larghezza di banda. Utilizza un'architettura pub/sub (pubblicazione/sottoscrizione), in cui i dispositivi comunicano attraverso un broker centrale che si occupa della distribuzione dei messaggi. I dispositivi publisher inviano messaggi su argomenti specifici (topic), e il broker li inoltra ai dispositivi subscriber che sono iscritti a quei topic.

Grazie alla sua capacità di garantire bassa latenza, affidabilità e un utilizzo ridotto della larghezza di banda, MQTT è ampiamente utilizzato in ambiti come l'Internet of Things (IoT), le reti di sensori, l'automazione industriale e, in particolare, nei sistemi medici (IoMT). La sua architettura è perfetta per ambienti che richiedono monitoraggio in tempo reale e gestione di dispositivi

distribuiti geograficamente. Inoltre, offre diverse opzioni di qualità del servizio (QoS), assicurando la consegna dei messaggi anche in caso di disconnessioni temporanee, mantenendo i messaggi in coda fino alla riconnessione dei dispositivi.

2 Metodi

Per la rilevazione di anomalie all'interno dell'infrastruttura ospedaliera, è stato sviluppato un sistema di analisi statistica del traffico basato sull'output fornito da ndpiReader.

2.1 Classificazione del traffico di rete

Come passaggio preliminare, il traffico di rete è stato classificato utilizzando nDPI, consentendo di isolare con precisione il traffico basato sul protocollo MQTT, permettendo un'analisi approfondita delle principali caratteristiche dei flussi di dati generati dai dispositivi IoT presenti nell'infrastruttura ospedaliera.

2.2 Metodologia per il Calcolo delle Statistiche Generiche

Per analizzare le caratteristiche dei flussi di rete, è stata adottata una metodologia strutturata che comprende il calcolo di diverse statistiche generiche, finalizzate a fornire una comprensione approfondita del comportamento del traffico. Questa metodologia si articola in diversi punti chiave:

- Quantificazione del Traffico
  - Totale dei Byte e dei Pacchetti: Sono stati calcolati i totali dei byte e dei pacchetti trasferiti nelle direzioni server-to-client e client-to-server, nonché i totali complessivi. Questi calcoli aiutano a quantificare il volume di dati e il numero di pacchetti scambiati nella rete.
- Analisi del Goodput
  - Byte di Goodput Totali: È stata effettuata la somma dei byte di goodput, che rappresentano i dati utili trasferiti senza considerare eventuali overhead o traffico non utile. Questo calcolo fornisce un'indicazione del throughput effettivo della rete.
- Valutazione dei Tempi
  - Durata Media dei Flussi: È stata calcolata la durata media tra i pacchetti per analizzare la frequenza e la distribuzione temporale del traffico. Queste informazioni sono cruciali per comprendere i pattern di comunicazione nella rete.
- Distribuzione dei Pacchetti e della Durata dei Flussi
  - Lunghezza dei Pacchetti: È stata analizzata la distribuzione della lunghezza dei pacchetti per identificare i pattern di dimensione dei dati scambiati. Questo aiuta a capire la tipologia di traffico prevalente.
  - Durata dei Flussi: È stata valutata la distribuzione della durata dei flussi per osservare come il tempo di vita dei flussi varia, il che può fornire indicazioni sui tipi di comunicazioni prevalenti nella rete.
- Identificazione dei Top Talkers
  - Top Talkers: Sono stati individuati gli IP principali responsabili della maggior parte del traffico, fornendo una vista sui principali attori nella rete e potenzialmente evidenziando fonti di traffico elevato.
- Distribuzione del Traffico per Protocollo
  - Protocollo e Traffico: È stata analizzata la distribuzione del traffico per protocollo, offrendo un quadro della percentuale di traffico generata da ciascun protocollo rispetto al totale. Questo aiuta a comprendere l'uso dei protocolli nella rete.
- Statistiche dell'Inter-Arrival Time (IAT)

– Minimo, Massimo e Media dell'IAT: Sono state calcolate le statistiche principali dell'Inter-Arrival Time per comprendere la variabilità e le tendenze nei tempi tra i pacchetti, fornendo un'indicazione della regolarità del traffico.

Questa metodologia complessiva consente di ottenere una visione dettagliata del comportamento e delle caratteristiche del traffico di rete, facilitando l'individuazione di eventuali anomalie o tendenze significative che potrebbero richiedere ulteriori indagini.

2.3 Metodologia per l'Identificazione degli Outlier

Per identificare gli outlier nei flussi di rete, abbiamo adottato un approccio basato sull'Interquartile Range (IQR), un indice di dispersione dei dati che misura di quanto i dati si allontanano da un valore centrale. Per garantire una valutazione completa delle anomalie, questo metodo è stato applicato a diverse metriche:

1. IQR del goodput dei bytes della cattura fornita rispetto ai flussi aggregati
2. IQR del goodput ratio dei flussi aggregati
3. IQR del ratio tra una flag specifica e il numero di pacchetti totali dei flussi aggregati

In tutte le metriche sviluppate, la metodologia rimane invariata:

- Calcoliamo il 25° percentile (Q1) e il 75° percentile (Q3) dei byte di goodput per tutti i flussi aggregati.
- Definiamo i limiti inferiori e superiori per gli outlier come

$$Q_1 - 1.5 \times IQR \quad e \quad Q_3 + 1.5 \times IQR$$

- Identifichiamo i flussi che cadono al di fuori di questi limiti come outlier.

3 Test

Per la validazione di MQTTAnalyzer, abbiamo condotto una serie di test utilizzando dataset forniti dal docente contenenti delle catture effettuate in un ambiente di rete simulato creato dai ricercatori dell'università del New Brunswick.

La Figura 3 illustra la topologia di rete utilizzata per gli esperimenti. In questa configurazione, i dispositivi Raspberry Pi e i relativi client malevoli sono collocati all'interno di una rete controllata, permettendo di simulare attacchi MQTT e raccogliere i dati per l'analisi.

L'infrastruttura di test è costituita da due dispositivi Raspberry Pi, ciascuno configurato con due client che operavano in qualità di agenti malevoli.

Device attaccanti		
Device	IP address	MAC address
RaspberryPI 1	192.168.137.48	dc:a6:32:dc:27:d5
RaspberryPI 1	192.168.137.156	dc:a6:32:dc:27:d5
RaspberryPI 2	192.168.137.180	dc:a6:32:c9:e4:d5
RaspberryPI 2	192.168.137.212	dc:a6:32:c9:e4:d5

Tabella I: La tabella mostra l'associazione tra i RaspberryPI utilizzati e i rispettivi indirizzi IP e MAC

I file di acquisizione (formati pcap e CSV), impiegati per l'analisi e necessari per la riproduzione dell'ambiente sperimentale, sono disponibili per il download al seguente link: Google Drive

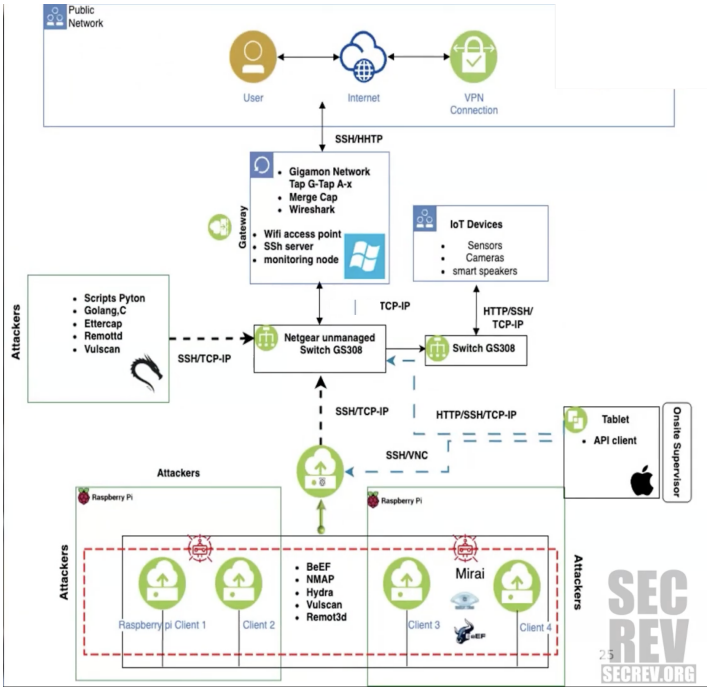


Figura 1: Topologia della rete. L'immagine proviene da una conferenza effettuata dall'istituto stesso recuperabile al link: From Profiling to Protection: Leveraging Datasets for Enhanced IoT Security

4 Risultati

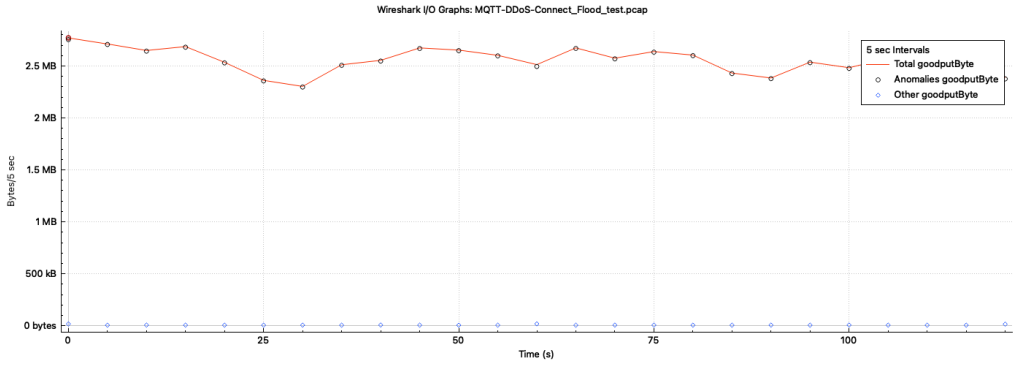
In questa sezione vengono analizzati i risultati ottenuti da MQTTAnalyzer sui vari dataset, confrontando i dati rilevati con le osservazioni effettuate attraverso Wireshark; a tal riguardo, poichè nei dataset forniti non è presente l'identità degli attacanti, la verifica dei risultati è stata effettuata utilizzando le sezioni "conversation" e "expert information dialog" di Wireshark e verificando che gli indirizzi segnalati dalla nostra applicazione avessero un comportamento anomalo paragonato al resto dei dispositivi presenti nella rete. Le anomalie rilevate sono state suddivise per feature, e i valori nella tabella II rappresentano il numero di anomalie identificate, con l'obiettivo di valutare la coerenza tra i due strumenti. A seguito del confronto abbiamo potuto constatare come gli indirizzi ip segnalati dalla nostra applicazione trovino un riscontro su Wireshark, riuscendo così a confermare gli obiettivi da noi prefissi.

Anomalie rilevate per feature					
Dataset	Feature1	Feature2	Feature3	Feature4	Wireshark
Benign	3	2	1	1	0
MQTT_DDoS_Connect	4	4	4	4	4
MQTT_DDoS_Publish	4	4	4	3	4
MQTT_DoS_Connect	3	1	1	1	1
MQTT_DoS_Publish	3	1	1	0	1
MQTT_malformed_data	3	1	1	1	1

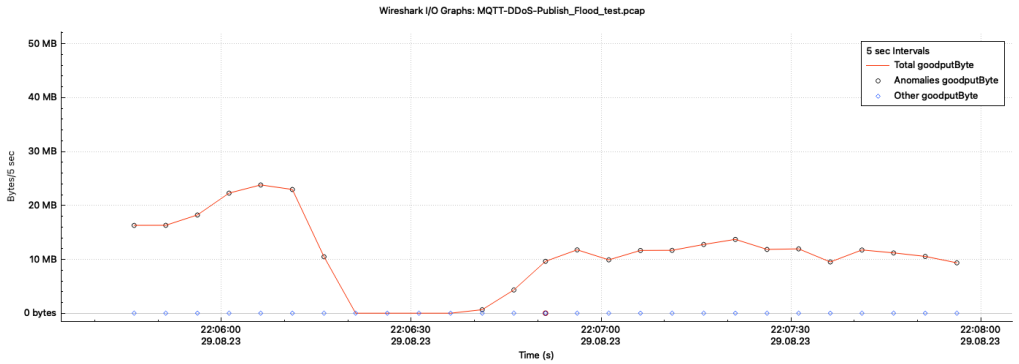
Tabella II: La tabella mostra il numero di anomalie rilevate da ciascuna feature in base al dataset usato, incluse quelle rilevabili tramite l'uso di Wireshark; di seguito il nome effettivo delle feature usate in ambiente di test: IQR_goodput_bytes, IQR_goodput_ratio, ACK_ratio, RST_ratio.

È stato riscontrato che alcune feature implementate possono rilevare falsi positivi (come è possibile osservare nella tabella II); per garantire una migliore affidabilità abbiamo quindi deciso di calcolare la similarità tra i risultati delle varie feature tramite l'utilizzo della similarità di Jaccard riducendo il numero di falsi positivi.

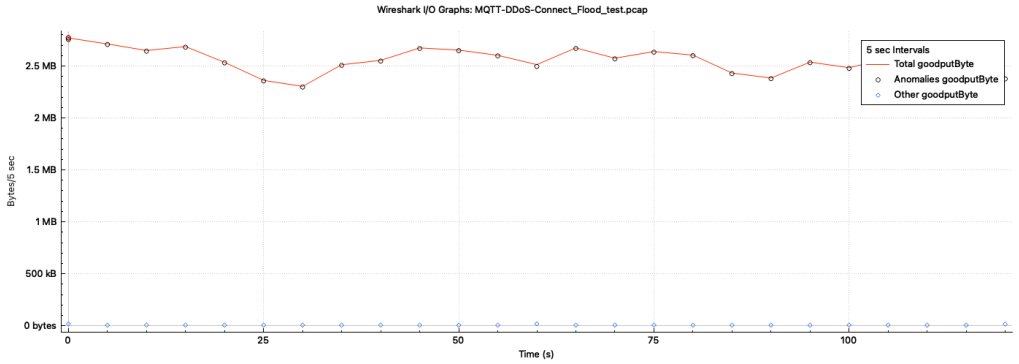
Di seguito vengono mostrati alcuni grafici generati tramite l'utilizzo di Wireshark in cui viene mostrato il quantitativo di bytes generati dagli indirizzi IP della rete; è possibile notare come il traffico generato dai device attaccanti (nel grafico identificato dal simbolo "o") comprenda la maggior parte del traffico totale (identificato dalla linea rossa) mentre il traffico generato dai device benigni sia molto più basso, indicando un evidente comportamento anomalo all'interno della rete concorde ai risultati ottenuti da MQTTAnalyzer.



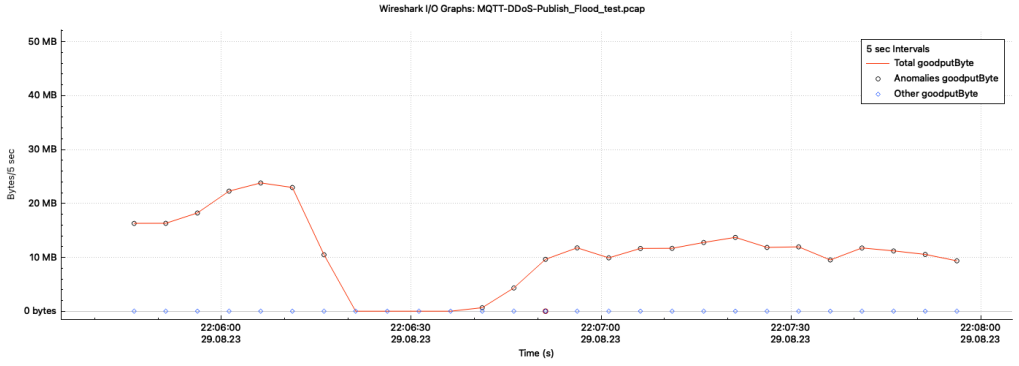
(a) MQTT DDoS Connect



(b) MQTT DDoS Publish



(c) MQTT DoS Connect



(d) MQTT DoS Publish