Data management and Storage IoT

Alessandro Schiavo

***Abstract*—L'Internet Of Thing è in costante diffusione data la facilità con la quale consente di raccogliere dati con dispositivi a basso costo di manutenzione ed costo energetico. Inoltre, la diffusione è data anche dalle basi solide in letteratura delle operazioni e dei sistemi per la raccolta, l'aggregazione, l'integrazione e visualizzazione del dato, qualunque sia la forma ed il contenuto.**

D'altro canto, l'impiego dei dispositivi su larga scala determina la necessità di approfondire gli aspetti citati onde evitare: sprechi di risorse, inutile esposizione a vulnerabilità note e scarsa tracciabilità dei dati raccolti. Quest'ultimo aspetto comprende: gestione delle periferiche offline e non di memorizzazione, rispetto delle politiche di privacy e sicurezza vigenti e ogni altro aspetto che riguardo la gestione dei dati a livello aziendale.

Nella seguente indagine sono approfonditi i temi di Data Management e Data Storage sintetizzando ed analizzando le proposte in letteratura più recenti. Sono quindi discussi temi come: gestione delle infrastrutture e dei dati, paradigmi più recenti infrastrutturali come Cloud, Edge, Fog Computing, tecniche di aggregazione e manipolazione dei dati per l'IoT.

I. Introduzione

L'industria dell'IoT impatta numerosi settori economici, dall'assistenza della vita privata delle persone, all'assistenza fornita alle organizzazioni per migliorare la qualità dei servizi offerti.

Le soluzioni IoT incentrate sulle persone, ad esempio, migliorano significativamente la routine quotidiana degli anziani e dei disabili, monitorano ed estraggono misurazioni vitali, oppure in altri contesti i veicoli connessi impediscono al conducente di deviare dalla traiettoria corretta. Gli innegabili vantaggi proposti dal paradigma IoT, tuttavia, si accompagnano a gravi carenze in termini di sicurezza.

Kashmir Hill in [1] cita un esempio in cui un negozio statunitense, Target, è stato in grado di rilevare la gravidanza delle donne grazie alla pubblicità e agli acquisti effettuati con la carta di credito e all'analisi dei loro acquisti di routine rispetto ai dati storici. Altri settori di applicazione sono sicuramente: sanità, produzione industriale, case intelligenti, città intelligenti e così via.

Le informazioni, sotto forma di dato aggregato, costituiscono valore aggiunto alle organizzazioni che ne fanno uso. In tale ambito, sono di cospicuo rilievo le scelte infrastrutturali per la costituzione dei sistemi IoT, oltre alle tecnologie per la gestione in sicurezza dei dati raccolti. L'IoT abbraccia tutti i dispositivi in grado di raccogliere informazioni; ciò determina la costituzione di infrastrutture di componenti eterogenee che fanno un uso inteso della reti di trasmissione.

Più che in altri settori, nell'IoT sono richieste caratteristiche tecniche in grado di garantire interoperabilità, affidabilità, standardizzazione dei formati ed adeguati livelli di privacy e conformità dei dati.

Da una visione macroscopica dei sistemi architetturali proposti in letteratura e presenti in commercio, si definiscono quattro macro-livelli di stratificazione:

1) primo livello costituito dai sensori e/o dispositivi
2) livello di comunicazione di rete formato da uno o più protocolli e/o sistemi di rete
3) livello di pulizia ed elaborazione dati e di estrazione delle caratteristiche di interesse
4) livello di visualizzazione e reportistica dei dati

I processi sopra elencati amplificano il ruolo che i dati rivestono, difatti la cultura del *Data Management* e della *Data Governance* sono cresciute di pari passo all'impiego dell'IoT su larga scala: i dati rivestono anche il ruolo di asset col quale compiere scelte strategiche e, come spesso accade, di ottimizzazione dei processi produttivi aziendali.

La Data Management è l'insieme di tecniche e tecnologie adoperate durante l'intera fase di vita del dato ed in particolare comprende operazioni di aggregazione, integrazione, memorizzazione e tutte ciò che riguarda la manutenzione delle informazioni di interesse.

Dalla nuova visione dei dati prende piede anche il concetto di Data Governance: le organizzazioni stabiliscono flussi e responsabilità delle operazioni compiute sui dati, con l'ottica di standardizzare i processi applicati sui Big Data.

Le caratteristiche dei dati rilevati sono definite da molteplici fattori come: settore di applicazione, intervalli temporali di raccolta, dispositivi utilizzati, protocolli e sistemi di rete impiegati. La dinamicità della forma del dato costituisce uno dei punti cardine sui quali basare la scelta dell'infrastruttura di un sistema.

II. Lavori di ricerca affini

In letteratura sono presenti diverse indagini che analizzano la gestione di infrastrutture basate su sistemi IoT ed approfondiscono temi come gestione del dato e conseguenti. Le indagini selezionate offrono diversi punti di vista sul temi come Data Management e Storage per l'IoT.

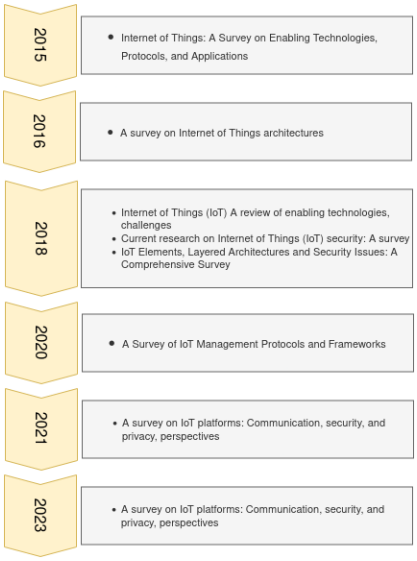

Fig. 1. Timeline delle indagini selezionate di riferimento per gli anni dal 2015 ad oggi

Guizani et al. in [2] analizza i concetti architetturali dei sistemi IoT confrontando caratteristiche come scalabilità e interoperabilità dei servizi forniti dai sistemi con tre e più strati architetturali. Segue l'analisi approfondita delle soluzioni proposte per affrontare le caratteristiche citate dei sistemi, oltre a proporre soluzioni per sicurezza, affidabilità e ottimizzazione delle prestazioni.

Ray et al. in [3] affronta le caratteristiche e le criticità delle tecnologie esistenti valorizzando i temi come manutenibilità dei sistemi e sicurezza degli accessi ai dati. Sono poi elencate le diverse proposte di ricerca come protocolli, proposte infrastrutturali e paradigmi emergenti. A seguire sono presentate delle linee guida per formalizzare gli ambiti di ricerca dove secondo lo studio sono presenti problemi infrastrutturali che non consentono ai sistemi IoT di essere applicati su larga scala.

Čolaković et al. in [4] confronta soluzioni architetturali analizzando i contributi più recenti. Sono descritte le esigenze applicative che comportano l'utilizzo di sistemi come: Cloud, Edge e Fog Computing ed approcci ibridi.

Hassan et al. in [5] propone uno studio sistematico delle tecnologie esistenti per l'integrazione, comunicazione e trasmissione dei dati al livello costituito da sensori e dispositivi hardware delle infrastrutture IoT.

Burhan et al. in [6] analizza gli aspetti critici di sicurezza e privacy sull'intero ciclo di vita dei dati raccolti focalizzando l'attenzione su protocolli e soluzioni infrastrutturali per garantire trasmissioni dei dati più sicure.

Sinche et al. in [7] studia i principali framework più consolidati negli ambiti di ricerca. Sono poi elencate le soluzioni dei provider presenti sul mercato, come quelle di Amazon Web Services ed IBM.

Babun et al. in [8] incentra lo studio su sicurezza del dato, risparmio ed ottimizzazione delle risorse disponibili. Si discute anche delle sfide che i database, relazionali e non, affrontano nei maggiori ambiti di applicazione dell'IoT, come scalabilità, efficienza, tempi di risposta e costi di archiviazione.

Ebenezer et al. in [9] si focalizza sulla valutazione delle piattaforme maggiormente in uso; evidenza la gestione corretta dell'intero ciclo di vita del dato elencando caratteristiche di diversi paradigmi architetturali come Cloud, Edge ed ibrido.

III. Edge e Cloud Computing: confronto e soluzioni proposte

A. Cloud Computing

Con il termine Cloud Computing si indicano una serie di strumenti forniti dai provider che astraggono la manutenzione, allocazione e l'inizializzazione delle risorse software e/o hardware. Il paradigma offre a diversi livelli di astrazione le risorse necessarie per lo sviluppo di prodotti di interesse. Sono di esempio i numerosi *data center* cloud che offrono risorse hardware a fronte di un costo che corrisponde al reale utilizzo delle risorse richieste, abbattendo ogni altro costo.

I tre principali livelli di astrazione delle risorse sono:

- Infrastructure as Service: offre capacità hardware come servizio e scalabili secondo le necessità, come Elastic Compute Cloud di Amazon Web Services. Le capacità offerte sono virtualizzate con interfacce standard integrabili con altri servizi
- Platform as Services: sono offerte piattaforme che astraggono la manutenzione, coprendo tutte le fasi del ciclo di vita del software e focalizzando l'attenzione in un'area di mercato ristretta
- Software as Services: comprende software gestito da uno o più provider ed offerto con la formula pay-per-use

Le risorse fornite dal Cloud Computing possono seguire l'andamento della richiesta risultando quindi altamente scalabili.

I dispositivi IoT hanno per natura un spazio di memorizzazione e risorse computazionali relativamente limitate al contesto; spostando le operazioni sui dati e la relativa memorizzazione sul cloud, questo garantisce, almeno concettualmente, risorse illimitate. Rimosso quindi il problema delle risorse da gestire, le organizzazioni che utilizzano in tal modo sistemi IoT riposizionano l'attenzione sui requisiti funzionali da fornire piuttosto che considerare anche i requisiti tecnici.

La figura seguente rappresenta l'infrastruttura dell'architettura di riferimento per IBM.

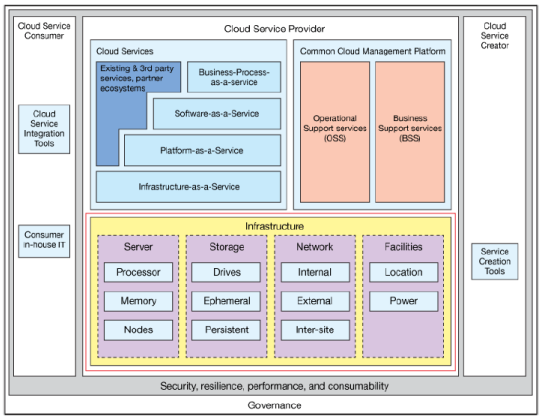

Fig. 2. infrastruttura dell'architettura di riferimento per IBM, Barillaud et al. in [10]

La figura mostra come diversi aspetti della manutenzione e della sicurezza delle risorse fornite dai servizi IaaS, PaaS e SaaS siano trasparenti per chi li utilizza. Inoltre, i provider possono definire uno o più servizi aggiuntivi di gestione e personalizzazione di aspetti tecnici e non funzionali, come: sicurezza, performance, resilienza e tutto ciò che riguarda la manutenzione dei dati.

B. Edge

L'edge computing è emerso come paradigma di trasmissione e raccolta dati provenienti da molteplici risorse, per contrastare noti problemi delle diverse soluzioni offerte dal Cloud Computing come congestione delle risorse e latenza di trasmissione. A differenza di quest'ultimo, le soluzioni Edge sono divise in livelli nei quali sono presenti uno o più tipologie di nodi che propagano i dati ricevuti ad altre risorse, applicandoci eventuali trasformazioni. I nodi costituiscono una fonte intermediaria dei dati ricevuti inizialmente dalle sorgenti e risultano utili per:

- rimozione di eventuali rumori sui dati grezzi
- memorizzazione parziale o totale dei dati ricevuti per garantire trasmissioni di dati efficienti
- operazioni di pulizia, trasformazione e caricamento dei dati provenienti dalle sorgenti con bassa latenza di ricezione e trasmissione
- fornire robustezza al sistema, con supporto alla duplicazione e ridondanza dei dati

In una normale applicazione di un sistema Cloud i dati sono trasmessi da multiple sorgenti a diverse distanze fisiche dalla destinazione, per poi essere ritrasmessi a queste quando le operazioni sui dati sono concluse. La conseguenza di questa tipologia di flusso è una congestione della rete questo può risultare come un ostacolo la realizzazione dei sistemi IoT safety critical e affini.

Le soluzioni architetturali dell'Edge Computing condividono uno schema tripartito nei livelli: il primo livello è costituito dalle sorgenti di dati (come attuatori, sensori) che sono posti nelle prossimità degli utenti finali così da esser facilmente installabili e tipicamente economici; un secondo livello costituito dai nodi interconnessi che redirigono i flussi di dati ad una o più destinazioni ed offrono caratteristiche come data caching, real time data processing e memorizzazione dei dati trasmessi; terzo ed ultimo livello, che tipicamente è costituito da risorse hardware maggiorate, per applicare operazioni tipiche de sistemi Cloud: big data mining, parallelizzazione dei processi sui dati, altro.

In letteratura esistono numerose ricerche sulla distribuzione dei nodi e sulla suddivisione di questi in gerarchie che considerano fattori come ruoli dei nodi stessi. Da una visione macroscopica della disposizione dei nodi, discendono i modelli:

1) modello gerarchico: i nodi sono fisicamente schierati a differenti distanze, con risorse allocate in funzione delle ruoli assegnati. Tra le differenti proposte in letteratura vi è Jararweh at al. [11] che propone una infrastruttura *Mobile Edge Computing* con nodi intermediari offrendo risorse computazionali e spazio di memorizzazione necessario agli utenti finali. Tong et al. [12] propone una struttura ad albero con nodi alle estremità delle rete e nodi regionali che aggregano partizioni delle rete dei nodi
2) modello software: in aggiunta al modello gerarchico sono proposte in letteratura modelli software per diminuire quanto possibile il grado di difficoltà della realizzazione di una rete IoT in diversi scenari dove l'ottimizzazione delle risorse riveste un ruolo cruciale. È di esempio la proposta di Manzalini e Crespi in [13]

IV. Data Aggregation

L'aggregazione dei dati risulta uno dei fattori cruciali per garantire la manutenibilità dei sistemi permettendo il risparmiarmio di memoria ed energia aggregando i dati secondo modelli predefiniti. Le ricerche che considerano la Data Aggregation si focalizzano sullo studio delle tipologie di aggregazione che risultano convenienti in base sia alla struttura del sistema che ai processi di analisi dei dati compiuti.

La Data Aggregation si accosta in maniera naturale al paradigma di **Fog Computing** che cattura i benefici del Cloud ed Edge Computing combinandoli in un'unica rete. In questa rete è aggiunto un livello intermediario costituito da nodi di rete con elevate capacità hardware che ricevono i dati dai nodi periferici e che trasmettono i dati ai servizi Cloud se necessario. Tra le tecniche di aggregazione più comuni risulta il WSN (Wireless Sensor Networks), le reti strutturate e non strutturate.

A. Reti

Gli algoritmi di data aggregation sono applicati sulla base della tipologia di rete dei nodi di trasmissione dei dati.

In uno schema gerarchico, solitamente ad albero, i nodi dell'ultimo livello sono le sorgenti ed è qui che risiedono i dati grezzi. Hoang et al in [14] propone un algoritmo per incrementare l'ottimizzazione ottenuta dalla Data Aggregation assegnando etichette ai nodi in base ad una probabilità che ne attesta la capacità di incrementare le prestazioni.

Lo schema più popolare è lo schema basato su cluster di nodi; i cluster raggruppano i nodi per caratteristiche simili e

i dati sono condivisi tra cluster solo quando sui dati di un cluster sono applicate le operazioni di riduzione.

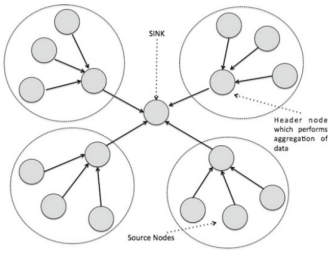

Fig. 3. Esempio di architettura basata su cluster, Yadav et al. in [15]

Esistono anche reti non strutturate nelle quali se non è presente un criterio di trasmissione, come priorità e caratteristiche dei nodi, le connessioni sono definite secondo algoritmi pseudocasuali.

Rezaeibagha et al. in [16] propone un meccanismo per collezionare e accumulare dati provenienti da dispositivi IoT indossabili utilizzano la crittografia omomorfa. La peculiarità dei meccanismo è che consente di effettuare operazioni di analisi dei dati direttamente dai dati crittografati, così da rendere superflue le operazioni di anonimizzazione e simili da applicare sui dati raccolti dagli utilizzatori dei dispositivi IoT.

Sanyal et al. in [17] propone un processo aggiuntivo da applicare prima delle operazioni di analisi. Il processo si propone di rimuovere le eventuali distorsioni presenti derivanti da molteplici casi di errore che possono intercorrente durante le fasi di rilevamento. Il processo è di tipo non supervisionato, quindi applicato senza alcuna informazione preliminare. La peculiarità dello schema di operazioni è di rimuovere le incertezze preservando le caratteristiche generali presenti sui dati nonché rimuovendo eventuali dati ridondanti o parziali.

V. Data Collection

La Data Collection comprende una serie di approcci utilizzati per ottenere consumi energetici ridotti e minor consumo delle risorse di trasmissione e di computazione. Nella figura di seguito sono elencati diversi approcci:

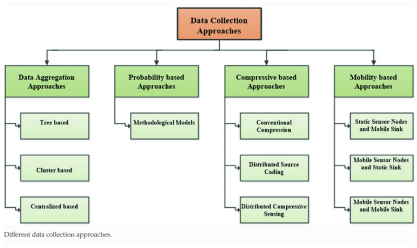

Fig. 4. Approcci Data Collection, Wala et al. in [18]

Wang et al. in [19] propone la metodologia *Approximate Data Gatehering* come metodologia efficiente per le numerose applicazione della WSN a causa delle limitate capacità di trasmissione e scarse capacità energetiche. La rete è divisa in cluster e per ognuno di questi si conduce una raccolta dei dati approssimata secondo parametri definiti per il cluster; con tali parametri è utilizzato un modello matematico per descrivere le correlazioni tra dati attuali e storici.

A. Soluzioni proposte per Data Aggregation e Data Collection

Le caratteristiche fondanti dei sistemi IoT ricadono nei benefici sia delle proposte Edge che Cloud Computing. L'IoT richiede in primis capacità computazionali elevate e grandi capacità di memoria a fronte di risposte a bassa latenza. Le architetture Edge offrono una buona tolleranza alle elevate richieste computazionali e ottimi tempi di risposta a fronte di una gestione più complessa della memoria e sincronizzazione di questa tra i diversi nodi.

D'altro canto, il ramo dell'Edge Computing trae significativamente vantaggio dall'applicazione di componenti dell'IoT rendendo la rete distribuita e dinamica. Dato il numero crescente dei dispositivi IoT utilizzati, l'Edge Computing e l'Internet Of Thing sembrano destinati a diventare inseparabili.

Le tempistiche di trasmissione e ricezione sono elemento fondante dei sistemi che utilizzano dispositivi IoT; è di esempio il progetto di ricerca presentato da Microsoft [20] che propone la realizzazione di un sistema in grado di predire congestione di flussi su rete distribuite come flussi di traffico. Il sistema inoltre segue ottimizzazioni sui costi, sulla manutenzione e sui tempi di risposta.

Lo schema e la distribuzione dei nodi e al bilanciamento del carico nelle reti Edge costituiscono una fattore determinante per l'ottimizzazione delle risorse utilizzate; Barcelo et al. in [21] propone un framework che tratta le reti di nodi come un problema di flusso riducendo con un fattore di al massimo 80% le risorse considerate.

Nei sistemi IoT l'archiviazione dei dati è resa complessa dalla quantità di dati da memorizzare. Con i concetti cardine dei sistemi di Edge Computing i nodi possono utilizzare parte della propria memoria per salvaguardare le prestazioni della rete in caso di errore. Ad esempio, parte della memoria può sfruttare i concetti di ridondanza e disponibilità per garantire il corretto funzionamento della rete anche in casi di errori a livello di nodi che a livello di memorie. Uno studio condotto da [22] mostra come meno del 10% degli errori dura più di 10 minuti. Ancora lo studio, condotto da Ford et al. in [22] propone un algoritmo per calcolare statisticamente il tempo medio al fallimento (MTTF) e calcolare il tempo medio alla risoluzione dell'errore espressa in probabilità.

In caso di un elevato numero di richieste localizzate su una regione della rete, risulta utile assegnare punteggi di priorità alle richieste per migliorare i tempi di risposta. La scala di priorità può essere assegnata in base a diversi fattori come: tempo, risorse necessarie e tipologia di richiesta. You et al. in [23] propone il calcolo dell'etichetta in base alle risorse energia e computazione del dispositivo locale e dal guadagno ottenuto dalla richiesta per l'intera rete.

Il modello Smart City si propone ottimizzare l'utilizzo delle risorse pubbliche ed incrementare la qualità della vita delle città. A riguardo Sapienza et al. in [24] propone uno studio sulla prevenzione di eventi anormali o critici (attacco terroristico, disastri naturali, ecc.). La proposta di Zhang et al.

[25] comprende il framework EVAPS (Edge Video Analysis for Public Safety) e con questo sono utilizzati sia i nodi della rete che un sistema Cloud Computing integrato così da eliminare eventuali dati non utili al processo di prevenzione e incremento della sicurezza delle strade urbane.

VI. Data analytics

L'analisi delle caratteristiche di interesse dei dati ricavati dalle operazioni ETL/ELT richiede approcci matematici e statistici, dato il numero cospicuo di dati da elaborare. La fase di analisi analizza, estrae, visualizza e compie inferenza sui dati disponibili con modelli matematici predittivi.

Sono elencati di seguito i livelli di analisi più comunemente utilizzati:

- livello di memoria: applicata quando la dimensione del pacchetto di dati è inferiore alla memoria disponibile, facilmente utilizzabile per calcoli in tempo reale. Un esempio è MongoDB.
- livello BI: viene adottata quando la dimensione dei dati è maggiore della memoria disponibile. Con gli strumenti attualmente presenti sul mercato è alla portata di tutti ricavare facilmente dai dati grezzi le principali correlazioni tra dati, fornendo anche una inferenza approssimata sui nuovi dati.
- analisi massiva: quando la mole di dati è superiore alle capacità dei comuni database e capacità dei prodotti BI, allora i dati contenuti sono accumulati per poi essere analizzati in un secondo momento. Solitamente sono applicati numerose funzionalità di pulizia e trasformazione dei dati.

A. Tecniche

Di seguito sono riportati i principali algoritmi applicati a seguito di fasi di trasformazione del dato, eseguite con strumenti come: MapReduce, Spark, Splunk e Skytree in grado di analizzare i grandi set di dati IoT.

- Reti bayesiane: Le reti bayesiane sono efficienti per analizzare strutture di dati complesse rivelate attraverso i big data piuttosto che i tradizionali formati di dati strutturati. Queste reti sono grafi aciclici diretti. L'analisi dei modelli di dati e la creazione di gruppi vengono eseguite in modo efficiente utilizzando SVM, che è anche un approccio di classificazione per l'analisi dei big data
- KNN: Allo stesso modo, KNN è tipicamente progettato per fornire meccanismi efficienti per trovare modelli nascosti da insiemi di big data, in modo tale che gli oggetti recuperati siano simili alla categoria predefinita.
- Clustering : Il clustering è un'altra tecnica di data mining utilizzata come metodo di analisi dei big data. Contrariamente alla classificazione, il clustering utilizza un approccio di apprendimento non supervisionato e crea gruppi per determinati oggetti in base alle loro caratteristiche distintive significative. L'approccio del clustering gerarchico continua a combinare piccoli cluster di oggetti dati per formare un albero gerarchico e creare cluster agglomerativi. I cluster divisivi vengono creati in modo opposto dividendo un singolo cluster che contiene tutti gli oggetti dati in cluster appropriati più piccoli.
- Analisi predittiva: l'analisi predittiva utilizza dati storici, noti come dati di training, per determinare i risultati come tendenze o comportamento nei dati.

B. Visualizzazione del dato

La visualizzazione è un aspetto di rilevanza nell'utilizzo dei Big Data raccolti dai sistemi IoT per i quali il tempo di risposta è un fattore determinante. Le più comuni soluzioni di cloud computing offrono strumenti all'avanguardia che semplificano la rappresentazione dei dati multidimensionali raffigurandone le principali caratteristiche. Le possibili difficoltà sono:

- rumore visivo: può capitare due più entità da rappresentare siano correlate semanticamente/ sintatticamente e questo facilmente comporta una errata percezione agli utenti lasciando intendere correlazioni tra dati non correlati;
- perdita di informazioni: l'applicazione di metodi di riduzione ai dataset può causare perdita di informazioni rilevanti;
- osservazione di immagini e grafici complessi: gli strumenti di visualizzazione dei dati affrontano i problemi relativi alle proporzioni, alla risoluzione degli oggetti rappresentati, nonché i limiti della percezione fisica;
- dati che cambiano frequentemente: l'utilizzatore finale può non notare cambiamenti repentini e poco impattanti

Dunque la rappresentazione del dato risulta un fattore determinante per trasmettere il valore dei caratteristiche selezionata. Per semplificare e raggruppare le caratteristiche interessanti sono applicate tecniche di riduzione, aggregazione e di riduzione delle dimensioni. Wang et. al in [26] propone il metodo DGE per separare le caratteristiche multidimensionali. Zhong et. al in [27] propone un metodo per visualizzare in real-time i Big Data memorizzati nel cloud.

C. Strumenti per la Data Visualization

Esistono in commercio numerosi e vari strumenti per la Data Visualization; scegliere correttamente gli strumenti consente di sfruttare al meglio le caratteristiche dei dati.
I Tableau sono uno strumento interattivo, di facile utilizzo e supporta l'aggiornamento automatico delle dashboard.
Kibana risulta adeguato per i casi di analisi avanzata sui dati, come esplorazione delle correlazioni tra diverse osservazioni ed applicazione di algoritmi di machine learning.

VII. Data Integration

I dati provenienti da diversi sensori e dispositivi sono aggregati per fornire una chiara visione generale dell'ecosistema. La data integration è un processo che comprende varie fasi di pulizia, organizzazione e aggregazione dei dati cosicché il fruitore finale sia in grado dare una interpretazione ai dati con una visione generale sull'intero sistema. Un tipico flusso di operazioni applicate in questa fase sono:

- data extraction: raccolta dei dati ed estrazione delle feature di interesse adoperando algoritmi studiati per la tipologia di problema e per la famiglia di dati in esame.

In questa fase può avvenire una parziale o totale perdita di informazioni rilevanti per il sistema
- data trasformation: Le trasformazioni applicate sui dati sono messe all'opera per ridurre la presenza di dati irrilevanti o dati parzialmente mancanti. Le trasformazioni includono operazioni di merge, normalizzazione e l'applicazione di operazioni ETL o ELT (extract , transform and load). Il secondo risulta particolarmente utile per i sistemi cloud, poiché carica i dati grezzi direttamente nell'archivio di massa centralizzato, preparando in tal modo i dati a trasformazioni da applicare in un secondo momento
- data mapping: processo di integrazione dei dati. Questa fase può avvenire anche nella Data Trasformation. I dati provenienti da diversi sistemi possono avere nomenclature e strutture dati diverse, generando molto spesso dati ridondati o superflui oltre a sovrapposizioni dello spazio dei nomi
- data loading: il caricamento dei dati su altre risorse di supporto può costituire un'operazione dispendiosa; per questo motivo sono utilizzate tecniche come caricamento batch per ridurre la banda di trasmissione utilizzata
- data validation: nell'archivio finale i dati sono convalidati per verificarne l'accuratezza e la completezza controllando che soddisfino criteri predefiniti o regole aziendali

VIII. Data Retention Policy

La conservazione dei dati e la conseguente gestione è una questione cruciale in tutte le organizzazioni che operano sui dati, specialmente per quelle che ne utilizzano ingenti quantità e che contengono informazioni delicate come la salute dei pazienti ospedalieri. La Data Retention Policy che un'azienda segue è costituita da normative aziendali e da normative di settore o territoriali, come la GDPR per gli stati membri dell'Unione Europea.

In base all'infrastruttura dell'architettura IoT, la cancellazione dei dati può risultare più o meno costosa e difficile da applicare e può dipendere dai tempi di conservazione e da altri fattori. Nei servizi offerti da provider terzi che automatizzano la gestione delle risorse di memorizzazione, la cancellazione dei dati risulta trasparente per l'organizzazione fruitrice. D'altro canto, se l'infrastruttura di base predispone di spazi di memorizzazione, come nel caso dell'Edge computing, la gestione dei dati secondo le normative in vigore può risultare ardua soprattutto se sono presenti spazi di memorizzazione ridondanti o offline (che è pratica comune per assicurarsi una maggiore tolleranza dell'errore).

Diène at al. [28] propone di crittografare ogni file con una chiave univoca e di utilizzarla per controllare la durata del file in memoria; gestendo in maniera centralizzata le chiavi, il sistema può sapere in ogni momento lo stato dei file sia su dispositivi online che offline.

IX. Sicurezza e Privacy

La natura complessa e interconnessa dell'IoT lo rende vulnerabile. I dati raccolti attraverso i sistemi IoT devono avere buoni meccanismi di controllo della privacy, ciò è possibile solo se sono applicate adeguate strategie di mitigazione del rischio.

Spesso la limitata conoscenza del dominio o di prospettive di rischio inesplorate comporta una totale o parziale mancanza di una valutazione periodica di esposizione al rischio delle organizzazioni che fanno uso dei sistemi IoT, ma che è resa necessaria dall'elevata dinamicità delle infrastrutture che tipicamente caratterizzano tali sistemi.

Gli ambiti di privacy e sicurezza dei dati nell'IoT assumono un ruolo cruciale non solo per il singolo individuo o l'organizzazione che subisce il danno, ma per interi settori commerciali. Questa criticità è dovuta in primo luogo dall'eterogeneità dei dispositivi IoT. Di seguito sono elencati tre requisiti fondamentali di sicurezza dei dati:
- riservatezza: con riferimento alla protezione dei dati da accessi non autorizzati
- integrità: con riferimento alla protezione dei dati da modifiche non autorizzate
- disponibilità: con riferimento alla garanzia che i dati siano sempre disponibili

A questi si aggiunge la privacy che estende il concetto di riservatezza dei dati. Per privacy si intende non solo il concetto di riservatezza, ma anche il rispetto delle preferenze della persona e il rispetto delle normative legali. Oltre alla normale difficoltà di eseguire gli standard di privacy da ogni strumento impiegato, il diretto interessato potrebbe modificare nel tempo il suo parere riguardo la condivisione dei dati stessi.

A. Vulnerabilità e rischi legati ad utilizzi impropri dell'IoT

I sistemi IoT sono caratterizzati da elevata granularità e dinamicità e questo spesso comporta una esposizione elevata ai vulnerabilità su sicurezza e privacy.

Come da un lato può sussistere l'inconsapevolezza delle aziende IT, d'altro canto applicare le tecniche su sistemi IoT può risultare spesso difficile o oneroso date le piattaforme hardware limitanti. Diverse difficoltà tecniche dunque, tra cui capacità di archiviazione, potenza e calcolo limitate, rendono difficile affrontare vari requisiti di sicurezza dell'IoT.

B. Principali vulnerabilità

Sono elencate le più comuni vulnerabilità riscontrare nelle tecnologie e nei sistemi attualmente adoperati:
- *Energia insufficiente*: I dispositivi IoT tipicamente hanno energia limitata e non possiedono necessariamente la tecnologia o i meccanismi per rinnovarla automaticamente. Un utente malintenzionato potrebbe consumare l'energia immagazzinata generando un'ondata di messaggi legittimi o danneggiati, rendendo i dispositivi non disponibili per processi o utenti validi
- *Autenticazione inadeguata*: I vincoli unici nel contesto del paradigma IoT, come l'energia limitata e la potenza computazionale, mettono alla prova l'implementazione di meccanismi di autenticazione complessi
- *Crittografia impropria*: La protezione dei dati è di fondamentale importanza negli ambiti IoT, in particolare quelli che operano in mercati con dati vulnerabili e delicati

- *Porte aperte non necessarie*: I vari dispositivi IoT hanno porte aperte inutilmente durante l'esecuzione di servizi vulnerabili, consentendo a un utente malintenzionato di connettersi e sfruttare numerose vulnerabilità
- *Controllo degli accessi insufficiente*: Una forte gestione delle credenziali dovrebbe proteggere i dispositivi e i dati IoT da accessi non autorizzati.
- *Capacità di gestione delle patch inadeguate*: I sistemi operativi IoT e il firmware/software incorporato dovrebbero essere adeguatamente *patchati* per ridurre al minimo i vettori di attacco e aumentare le loro capacità funzionali. Tuttavia, numerosi casi riportano che molti produttori non mantengono regolarmente le *patch di sicurezza* o non dispongono di meccanismi automatizzati di aggiornamento.

Mahomoud et al. in [29] presenta un sondaggio sullo stato dell'arte dei sistemi di sicurezza ed elenca una serie di preoccupazioni riguardo le vulnerabilità più diffuse, tra cui la scarsa sicurezza fisica dei dispositivi IoT.

Furfaro et al. in [30] valuta le correnti soluzioni di sicurezza informatica utilizzando un approccio basato su ambienti virtuali.

C. Vulnerabilità per livelli

Di seguito sono divise le vulnerabilità per i tre macro livelli che costituiscono le infrastrutture IoT.

- *Vulnerabilità basate sui dispositivi*: Poiché un gran numero di dispositivi IoT funziona correttamente anche senza l'impiego di sistemi di sicurezza e metodologie di resistenza alla manomissione assenti o limitate, un utente malintenzionato potrebbe trarre vantaggio dall'accesso fisico a un dispositivo per causare danni significativi.
- *Vulnerabilità basate sulla rete*: Una serie di attività di ricerca hanno affrontato le vulnerabilità specifiche dell'IoT causate da debolezze della rete o del protocollo. Ad esempio, il protocollo ZigBee [31] sviluppato per reti di controllo e sensori wireless a bassa velocità/bassa potenza, consente i dispositivi che ne fanno uso di stabilire comunicazioni sicure implementando chiavi crittografiche simmetriche. In questo contesto, Vidgren et al. [32] ha illustrato come un avversario potrebbe compromettere i dispositivi IoT abilitati per ZigBee. Sebbene sia possibile la pre-installazione delle chiavi su ciascun dispositivo per una determinata modalità di sicurezza, in realtà le chiavi vengono trasmesse non crittografate.
- *Altre vulnerabilità*: Altre vulnerabilità sono spesso: manipolazione delle funzionalità con energia insufficiente del dispositivo, capacità di gestione delle patch inadeguate, controllo degli accessi insufficiente

D. Encryption

I device IoT sono in genere disposti di limitate risorse hardware che non consentono una normale applicazione degli algoritmi di cifratura e de-cifratura in relazione ad un utilizzo data intensive. Per tale scopo sul mercato sono disponibili versioni di microchip predisposte con un chip dedicato alle funzioni di cifratura. In letteratura sono presentate diverse forme efficienti dei più comuni algoritmi come [33] semplifica il processo di generazione delle chiavi AES. Un interessante punto di vista e' proposto da Hatzivasilis et al. [34] con la comparazione di 52 chip.

La cifratura dei dati non impedisce a malintenzionati di sfruttare l'accesso fisico al dispositivo per ottenere informazioni riservate; di esempio sono gli attacchi che sfruttano la manipolazione di risorse come corrente, campi elettromagnetici per dispositivi come i RFID ed ogni altro attacco applicabile direttamente sul sistema fisico.

E. Relazione tra sicurezza e dati di input

Le infrastrutture IoT sono per la più utilizzate per tracciare e memorizzare i dati provenienti dal mondo reale. Quella che sembra una normale funzionalità dei sistemi, in realtà può essere sfruttata da chi intende compromettere non solo tali infrastrutture, bensì la veridicità dei dati memorizzati. Ad esempio, un attaccante potrebbe inviare ad un sensore di un ospedale dei dati non veritieri sulla salute di un paziente attestando di conseguenza risultati incorretti dalla reportistica che l'infrastruttura genera.

Un immediato sistema di salvaguardia da tali tipologie di attacco e sicuramente mantenere i sistemi software costantemente aggiornati cosicché l'accesso ai singoli dispositivi non possa sfruttare vulnerabilità note. A tal proposito alcune case produttrici di componentistica offrono la possibilità di aggiornare il componente da remoto. Anche questa funzionalità, malauguratamente, può essere sfruttata per inviare versioni obsolete e con vulnerabilità conosciute, facilitando ancor più l'attaccante.

In letteratura sono presenti diversi elaborati a riguardo, come Ruckebush et al. [35] che propone aggiornamenti remoti parziali divisi in tre livelli di priorità.

F. Vulnerabilità delle comunicazioni multiple e simultanee

I protocolli di rete che per natura consentono comunicazioni simultanee tra diversi dispositivi sono tra le principali fonti di vulnerabilità, dove data la facilità di comunicare con multipli dispositivi simultaneamente permette all'attaccante di inviare con estrema facilita dati malevoli a multiple destinazioni. Spesso i livelli di sicurezza applicati per le comunicazioni all'interno della stessa sotto-rete non sono elevati poiché si presuppone che per accedervi siano oltrepassati diversi protocolli di sicurezza, cosa che ovviamente facilita gli attaccanti.

X. Scalability

Una ideale sistema IoT è in grado di incorporare qualsiasi dispositivo in grado di trasmettere dati, oltre alle proprie capacità hardware, alla tipologia di dati trasmessi e privando la fase di integrazione di ogni interazione con un essere umano. Oltre a tale capacità si intendono presenti una serie di requisiti non funzionali del sistema come scalabilità ed una adeguata soglia di fault tolerance.

Per archiviare le caratteristiche elencate non può mancare la capacità di scalabilità di un sistema IoT, sotto ogni punto

di vista. Con il termine scalabilità si intendono una serie di proprietà in grado di fornire dei servizi definiti oltre al numero di richieste ricevute dal sistema, delle risorse disponibili e alla presenza di eventuali errori fisici o software.

Il termine scalabilità quindi assume un significato macroscopico che racchiude in sé diversi requisiti a diversi livelli di dettaglio del sistema. Di seguito, con i termini *orizzontale* e *verticale* si enunciano le due più utilizzate nomenclature dei requisiti:

- scalabilità orizzontale: la capacità di distribuire il carico di lavoro utilizzando più o meno risorse hardware e/o software cosicché siano fornite le stesse prestazioni a fronte di carichi diversi
- scalabilità verticale: è la capacità di variare l'utilizzo delle risorse hardware e software in base alle reali necessità; eventualmente variando: supporti di elaborazione, memorie volatili, supporti di archiviazione e interfacce di rete. Inoltre, l'implementazione è più semplice, riduce i costi del software e viene mantenuta la compatibilità delle applicazioni. Come conseguenze del requisito è evidente una maggiore efficienza nell'utilizzo delle risorse che il sistema dispone

A. Tecniche per la scalabilità

Di seguito sono presentate diverse tecniche per fornire le proprietà di scalabilità a più livelli di sistema:

- Allocazione risorse automatizzato: i dispositivi devono disporre di strutture integrate con boot loader richiesti, chiavi di sicurezza e altre funzionalità necessarie che promuovano il processo di automazione quando un dispositivo situato in remoto si avvia per la prima volta; è di esempio lo studio condotto da Miorandi et al. in [36]
- Sistemi di controllo della pipeline di dati: con l'alta variabilità del numero di dispositivi che contribuiscono alla generazione e alla trasmissione dei dati, le pipeline di dati devono essere progettate in modo tale da poter gestire l'improvviso aumento o decremento delle richieste. E' di esempio lo studio di Tata et al. in [37]
- Sviluppo dell'architettura dei microservizi: I microservizi rappresentano un approccio architettonico contemporaneo in cui applicazioni complesse sono costituite da microprocessori individualistici che si diffondono tra loro con l'aiuto di API indipendenti dal linguaggio
- Adozione di più tecnologie di archiviazione dei dati: Un sistema IoT è composto da diverse applicazioni che richiedono tecniche diverse per la loro archiviazione e questo implica che per mantenere al meglio le diverse tipologie del dato sia utile fornire diversi sistemi di archiviazione, anziché utilizzarne uno per tutti
- Tolleranza agli errori: la scalabilità e la tolleranza agli errori sono strettamente collegate. Con la tolleranza agli errori si introduce anche la capacità di aumentare le risorse hardware e software per ovviare errori di entrambi. Allo stesso modo, con un processo che è in grado di scalare le sue risorse, anche dal punto di vista dell'eterogeneità, è più difficile che questo non sia in grado di gestire errori di un particolare hardware o software presenti.

Sarkar et al. in [38] propone un'architettura distribuita, Distributed Internet-like Architecture for Things (DIAT), in grado di assecondare la scalabilità e interoperabilità del sistema.

XI. Conclusioni

Dall'analisi condotta sullo stato dell'arte dei sistemi IoT emerge chiaramente come la gestione efficace dei dati attraverso i processi di raccolta, pulizia e memorizzazione, rappresentino aspetti fondamentali per garantire la qualità dei servizi offerti e sfruttare al meglio le risorse disponibili.

In commercio sono presenti numerosi provider che offrono servizi in grado di semplificare il processo di messa in opera dei sistemi che utilizzino dispositivi IoT. Con l'indagine condotta si evince come le criticità relative alla sicurezza dei dati raccolti non si possono considerare risolte, anche se sono disponibili soluzioni in larga scala per prevenire le vulnerabilità note. La presenza di vulnerabilità deriva maggiormente dall'eterogeneità dei supporti hardware, software e dalle non adeguate applicazioni dei sistemi di prevenzione; quest'ultimo aspetto è legato alla necessaria conoscenza approfondita del settore e dalle difficoltà tecniche di applicazione.

Temi come scalabilità, adozione di politiche di conservazione, integrazione e aggregazione dei dati sono emersi come elementi chiave per ottimizzare i costi di produzione e di gestione delle risorse. Sul mercato i servizi disponibili offrono in maniera trasparente alcune delle caratteristiche citate, con eventuali servizi di personalizzazione. Ciò non toglie che è compito delle organizzazioni assicurarsi l'adozione delle giuste tecniche per ottenere il massimo profitto e il giusto dispendio di risorse. Sono di esempio le proposte citate per l'aggregazione e integrazione dei dati per i sistemi distribuiti in cluster.

L'analisi dei dati e la conseguente visualizzazione, se applicate con criterio consentono di estrarre il reale valore dei dati raccolti. Come già citato, anche in questa fase sono presenti dettagli tecnici che le organizzazioni non devono trascurare onde evitare di estrarre informazioni fuorvianti o di scarso interesse.

References

[1] Kashmir Hill. "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did". In: *Welcome to The Not-So Private Parts where technology privacy collide* (2012).

[2] Ala Al-Fuqaha et al. "Internet of things: A survey on enabling technologies, protocols, and applications". In: *IEEE communications surveys & tutorials* 17.4 (2015), pp. 2347–2376.

[3] Partha Pratim Ray. "A survey on Internet of Things architectures". In: *Journal of King Saud University-Computer and Information Sciences* 30.3 (2018), pp. 291–319.

[4] Alem Čolaković and Mesud Hadžialić. "Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues". In: *Computer networks* 144 (2018), pp. 17–39.

[5] Wan Haslina Hassan et al. "Current research on Internet of Things (IoT) security: A survey". In: *Computer networks* 148 (2019), pp. 283–294.

[6] Muhammad Burhan et al. "IoT elements, layered architectures and security issues: A comprehensive survey". In: *sensors* 18.9 (2018), p. 2796.

[7] Soraya Sinche et al. "A survey of IoT management protocols and frameworks". In: *IEEE Communications Surveys & Tutorials* 22.2 (2019), pp. 1168–1190.

[8] Leonardo Babun et al. "A survey on IoT platforms: Communication, security, and privacy perspectives". In: *Computer Networks* 192 (2021), p. 108040.

[9] Calduwel Newton Jerome Oswald Ebenezer. "Data Management in IoT: A Detailed Survey". In: *International Journal of Information Technology Research and Applications* 2.2 (2023), pp. 18–32. DOI: 10.59461/ijitra.v2i2.49.

[10] Franck Barillaud, Chuck Calio, and J Jacobson. "IBM cloud technologies: How they all fit together". In: *IBM Corporation* (2015).

[11] Yaser Jararweh et al. "The future of mobile cloud computing: Integrating cloudlets and Mobile Edge Computing". In: *2016 23rd International Conference on Telecommunications (ICT)*. 2016, pp. 1–5. DOI: 10.1109/ICT.2016.7500486.

[12] Liang Tong, Yong Li, and Wei Gao. "A hierarchical edge cloud architecture for mobile computing". In: *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*. 2016, pp. 1–9. DOI: 10.1109/INFOCOM.2016.7524340.

[13] Antonio Manzalini and Noel Crespi. "An edge operating system enabling anything-as -a-service". In: *IEEE Communications Magazine* 54.3 (2016), pp. 62–67. DOI: 10.1109/MCOM.2016.7432173.

[14] Vipin Pal et al. "SCHS: Smart Cluster Head Selection Scheme for Clustering Algorithms in Wireless Sensor Networks". In: *Wireless Sensor Network* 4.11 (2012), p. 273. DOI: 10.4236/wsn.2012.411039. URL: https://doi.org/10.4236/wsn.2012.411039.

[15] Rajesh Kumar Yadav and Monica Gupta. "Data aggregation algorithms in IoT: An organized evaluation of the literature". In: *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE. 2020, pp. 300–304.

[16] Fatemeh Rezaeibagha et al. "Secure and Efficient Data Aggregation for IoT Monitoring Systems". In: *IEEE Internet of Things Journal* 8.10 (2021), pp. 8056–8063. DOI: 10.1109/JIOT.2020.3042204.

[17] Sunny Sanyal and Puning Zhang. "Improving Quality of Data: IoT Data Aggregation Using Device to Device Communications". In: *IEEE Access* 6 (2018), pp. 67830–67840. DOI: 10.1109/ACCESS.2018.2878640.

[18] Tanuj Wala, Narottam Chand, and Ajay K. Sharma. "Energy Efficient Data Collection in Smart Cities Using IoT". In: *Handbook of Wireless Sensor Networks: Issues and Challenges in Current Scenario's*. Ed. by Pradeep Kumar Singh et al. Cham: Springer International Publishing, 2020, pp. 632–654. ISBN: 978-3-030-40305-8. DOI: 10.1007/978-3-030-40305-8_30. URL: https://doi.org/10.1007/978-3-030-40305-8_30.

[19] Chao Wang et al. "Adaptive Approximate Data Collection for Wireless Sensor Networks". In: *IEEE Transactions on Parallel and Distributed Systems* 23.6 (2012), pp. 1004–1016. DOI: 10.1109/TPDS.2011.265.

[20] Ganesh Ananthanarayanan et al. "Real-Time Video Analytics: The Killer App for Edge Computing". In: *Computer* 50.10 (2017), pp. 58–67. DOI: 10.1109/MC.2017.3641638.

[21] Marc Barcelo et al. "IoT-Cloud Service Optimization in Next Generation Smart Environments". In: *IEEE Journal on Selected Areas in Communications* 34.12 (2016), pp. 4077–4090. DOI: 10.1109/JSAC.2016.2621398.

[22] Daniel Ford et al. "Availability in globally distributed storage systems". In: *9th USENIX Symposium on Operating Systems Design and Implementation (OSDI 10)*. 2010.

[23] Changsheng You et al. "Energy-Efficient Resource Allocation for Mobile-Edge Computation Offloading". In: *IEEE Transactions on Wireless Communications* 16.3 (2017), pp. 1397–1411. DOI: 10.1109/TWC.2016.2633522.

[24] Marco Sapienza et al. "Solving Critical Events through Mobile Edge Computing: An Approach for Smart Cities". In: *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*. 2016, pp. 1–5. DOI: 10.1109/SMARTCOMP.2016.7501719.

[25] Qingyang Zhang et al. "Demo Abstract: EVAPS: Edge Video Analysis for Public Safety". In: *2016 IEEE/ACM Symposium on Edge Computing (SEC)*. 2016, pp. 121–122. DOI: 10.1109/SEC.2016.30.

[26] Xiumei Wang et al. "A novel dimensionality reduction method with discriminative generalized eigendecomposition". In: *Neurocomputing* 173 (2016), pp. 163–171.

[27] Ray Y Zhong et al. "Visualization of RFID-enabled shopfloor logistics Big Data in Cloud Manufacturing". In: *The International Journal of Advanced Manufacturing Technology* 84 (2016), pp. 5–16.

[28] Bassirou Diène et al. "Data management mechanisms for IoT: architecture, challenges and solutions". In: *2020 5th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE. 2020, pp. 1–6.

[29] Rwan Mahmoud et al. "Internet of things (IoT) security: Current status, challenges and prospective measures". In: *2015 10th international conference for internet technology and secured transactions (ICITST)*. IEEE. 2015, pp. 336–341.

[30] Angelo Furfaro et al. "Using virtual environments for the assessment of cybersecurity issues in IoT scenarios". In: *Simulation Modelling Practice and Theory* 73 (2017), pp. 43–54.

[31] Shahin Farahani. *ZigBee wireless networks and transceivers*. newnes, 2011.

[32] Niko Vidgren et al. “Security threats in ZigBee-enabled systems: Vulnerability evaluation, practical experiments, countermeasures, and lessons learned”. In: *2013 46th Hawaii International Conference on System Sciences*. IEEE. 2013, pp. 5132–5138.

[33] Amir Moradi et al. “Pushing the Limits: A Very Compact and a Threshold Implementation of AES”. In: *Advances in Cryptology – EUROCRYPT 2011*. Ed. by Kenneth G. Paterson. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 69–88. ISBN: 978-3-642-20465-4.

[34] George Hatzivasilis et al. “A review of lightweight block ciphers”. In: *Journal of Cryptographic Engineering* 8.2 (2018), pp. 141–184. ISSN: 2190-8516. DOI: 10.1007/s13389-017-0160-y. URL: https://doi.org/10.1007/s13389-017-0160-y.

[35] Sanaz Rahimi Moosavi et al. “SEA: A Secure and Efficient Authentication and Authorization Architecture for IoT-Based Healthcare Using Smart Gateways”. In: *Procedia Computer Science* 52 (2015). The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015), pp. 452–459. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2015.05.013. URL: https://www.sciencedirect.com/science/article/pii/S1877050915008133.

[36] Daniele Miorandi et al. “Internet of things: Vision, applications and research challenges”. In: *Ad hoc networks* 10.7 (2012), pp. 1497–1516.

[37] Prateep Misra. “Build a scalable platform for high-performance IoT applications”. In: *TCS Exper. Certainty, Mumbai, India, Tech. Rep* (2016).

[38] Chayan Sarkar et al. “A scalable distributed architecture towards unifying IoT applications”. In: *2014 IEEE World Forum on Internet of Things (WF-IoT)*. 2014, pp. 508–513. DOI: 10.1109/WF-IoT.2014.6803220.