Exploring the Security Landscape of Open-Source IoT Software on GitHub

Zahra Hatefi
Mark van den Brand, Ivan Kurtev

Leiden University PhD Workgroup Meeting

EINDHOVEN UNIVERSITY OF TECHNOLOGY

- Introduction
- Study Design
  - Methodology
  - Inclusion and exclusion criteria
- Categories & characteristics
- Findings
  - IoT Systems, IoT security and IoT attacks
  - Security Aspects
  - Attacks
  - Communication protocols
- Conclusion and next steps

Introduction

Modeling security standards for Internet of Things using Domain specific language

❖ Address the conceptual foundation of IoT security and how it can be captured in a domain-specific language.

Main goal in the project

- Reasoning about security in IoT systems

Main steps

- Identifying IoT architecture and main component
- Conceptual model of IoT
- Identifying security aspects of IoT systems
- Place security in this context

Motivation

- We need concrete use cases to study how to model and reason about security in IoT systems and how security mechanism are implemented. looking into reality to see:
  - How software that makes secure IoT is developed
  - What kind of security problems they addressed
  - What kind of solutions they provided
  - How we can help developing this software more secure
- Look into the existing repos that solve security issues

❖ **There are no papers in software mining that focus on security in IoT projects.**

Introduction

- We explored:
  - The overarching landscape of open-source IoT software on GitHub
  - Focusing on:
    - Prevalent communication protocols
    - Commonly implemented security features
    - Common attacks
    - Vulnerabilities

Introduction

- We identify:
  - How many OSS IoT projects are existing on GitHub?
  - How many of them consider security concerns and attacks?
  - How many of them consider the vulnerabilities of their design?
  - What types of security properties and attacks are more common?
  - What kind of communication protocol did they use?

Study Design

Methodology

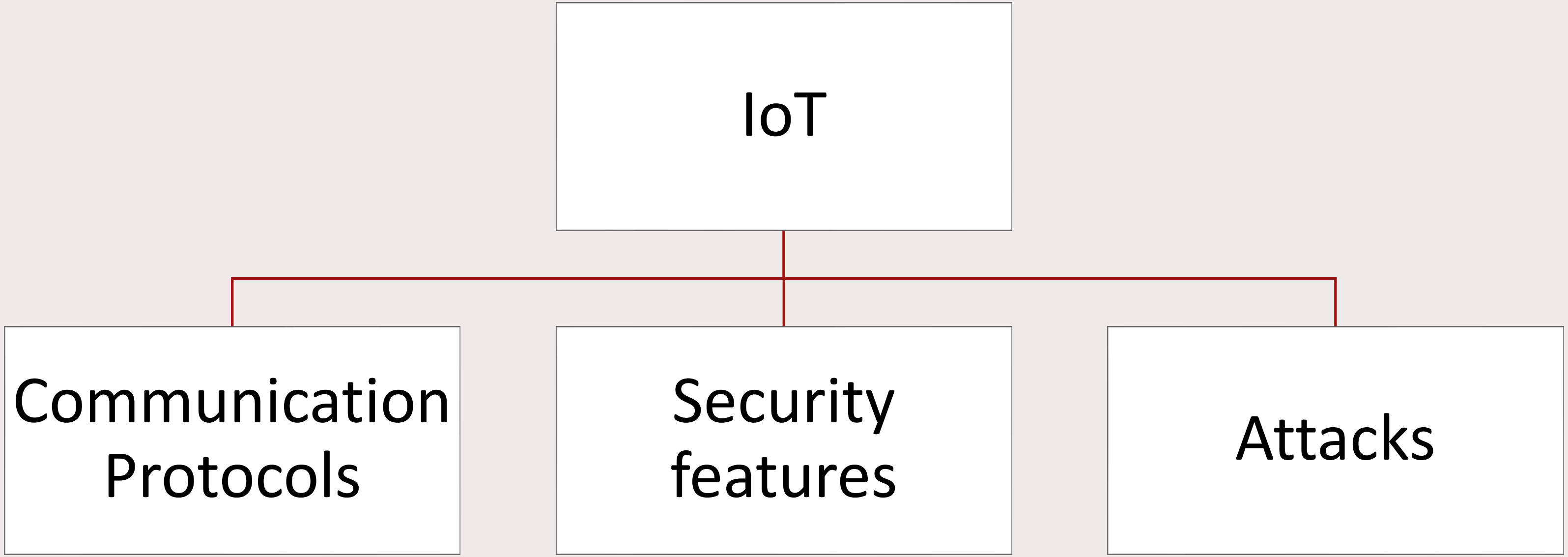
1. Identifying categories and characteristics
2. Identifying topics
3. Looking for resources on GitHub (by using PyGitHub)
4. Filtering the results and categorizing them

Inclusion and exclusion criteria

- Inclusion criteria:
  - The projects that have a size greater than 0.
  - The Repos, which was created between 2008 and 2024
- Exclusion criteria:
  - Repos that are not IoT or related to IoT
  - Repos which are not related to security aspects

Categories & characteristics

Categories & characteristics



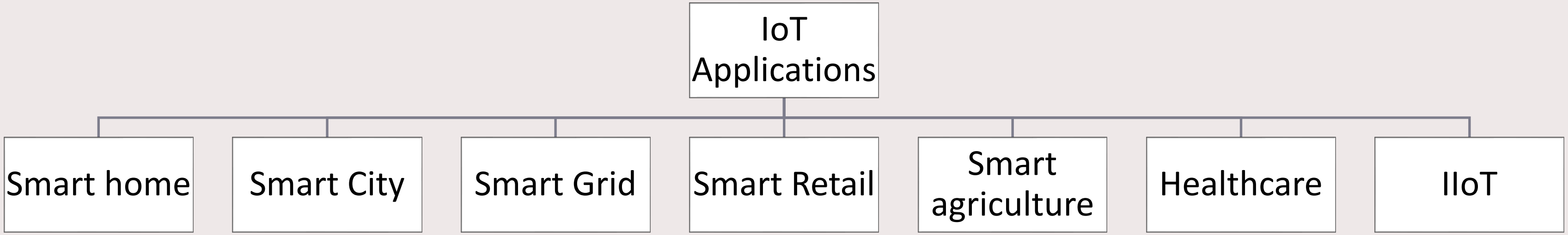
IoT

Communication Protocols

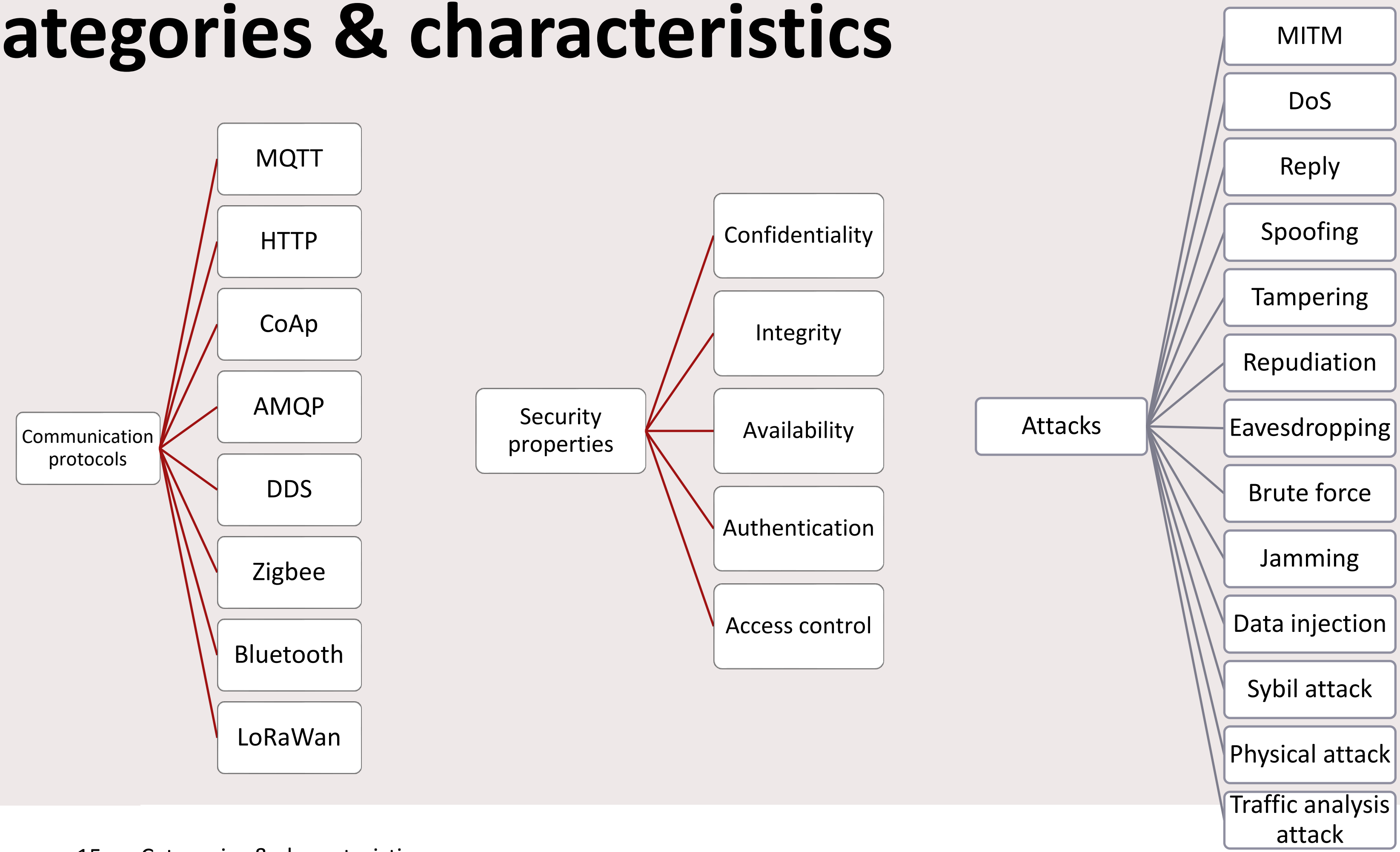
Security features

Attacks

Categories & characteristics



Categories & characteristics

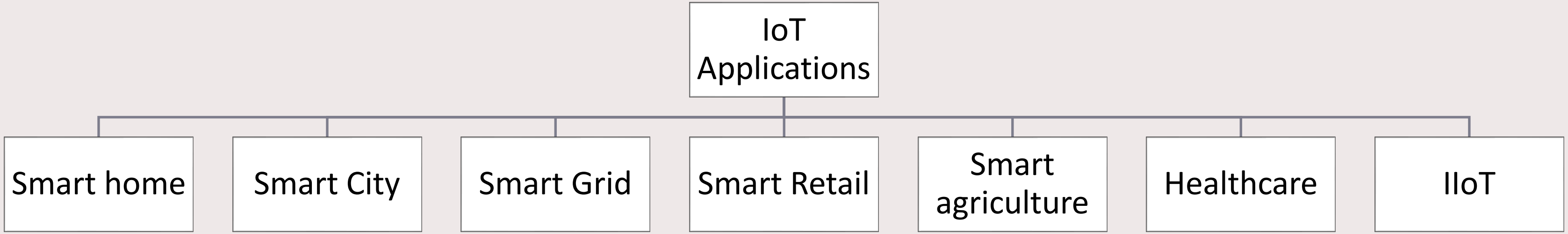


- Communication protocols
  - MQTT
  - HTTP
  - CoAp
  - AMQP
  - DDS
  - Zigbee
  - Bluetooth
  - LoRaWan
- Security properties
  - Confidentiality
  - Integrity
  - Availability
  - Authentication
  - Access control
- Attacks
  - MITM
  - DoS
  - Reply
  - Spoofing
  - Tampering
  - Repudiation
  - Eavesdropping
  - Brute force
  - Jamming
  - Data injection
  - Sybil attack
  - Physical attack
  - Traffic analysis attack

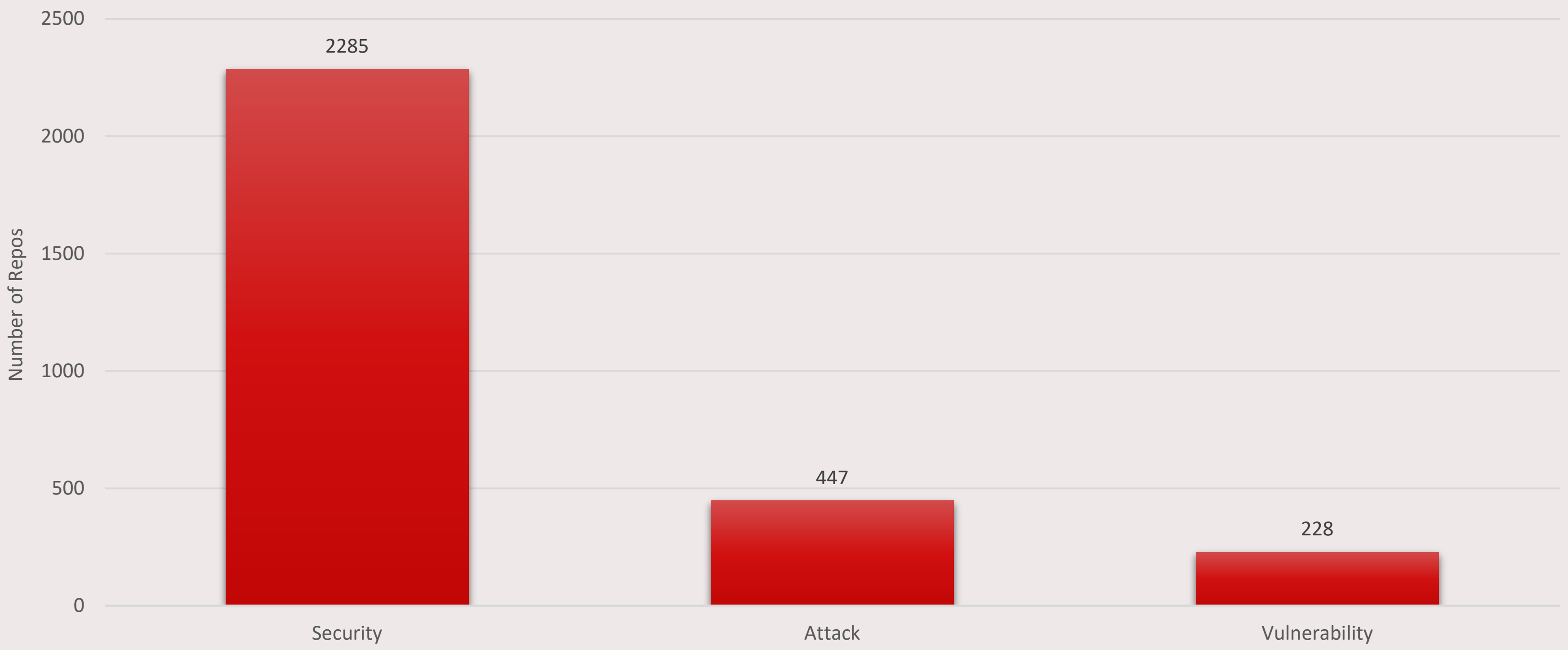
Findings

IoT repos

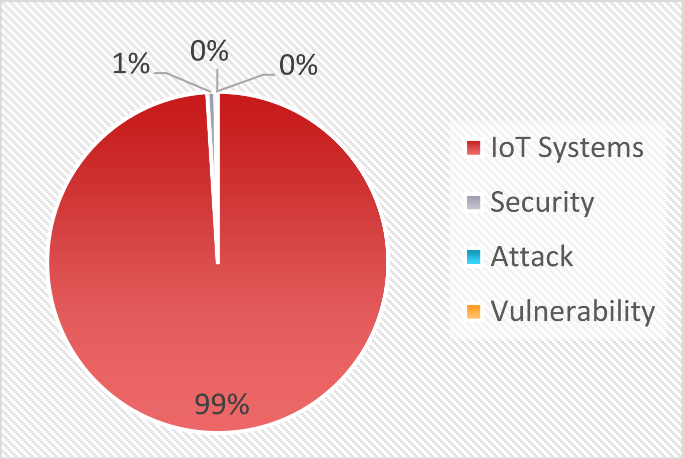
- We Identify ≈296978 IoT repositories on GitHub.



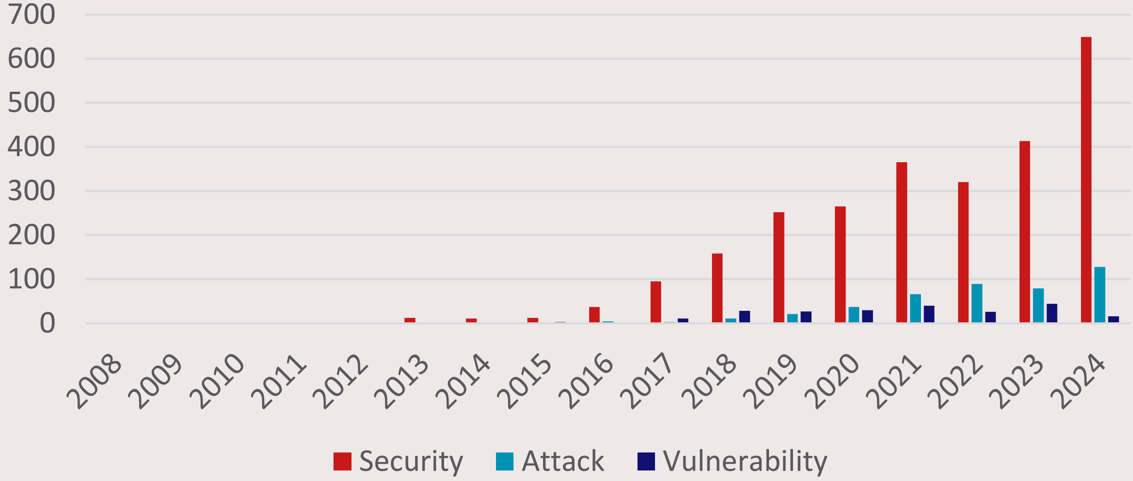
IoT systems and security, attack, and vulnerability among IoT repos



IoT systems and security, attack, and vulnerability among IoT repos

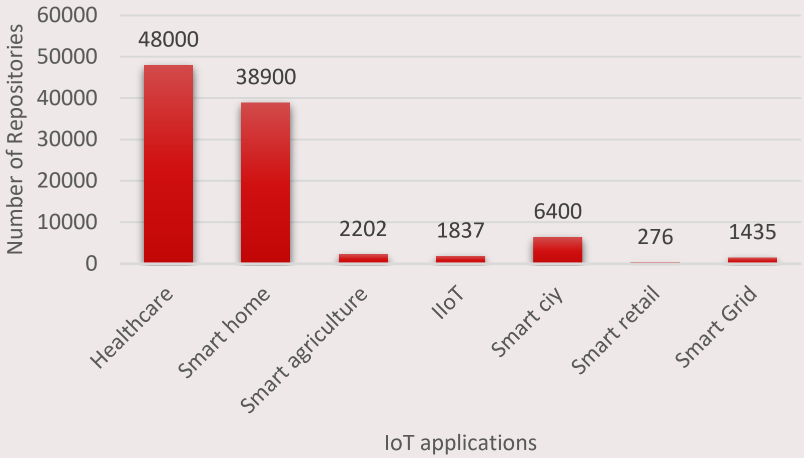


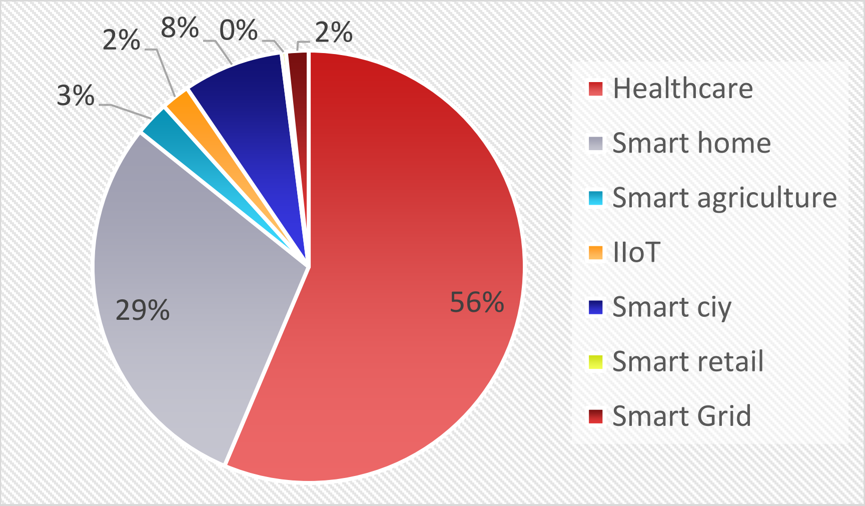
Temporal evolution of developed IoT repositories, which consider the security and attacks on GitHub, between 2008 and 2024



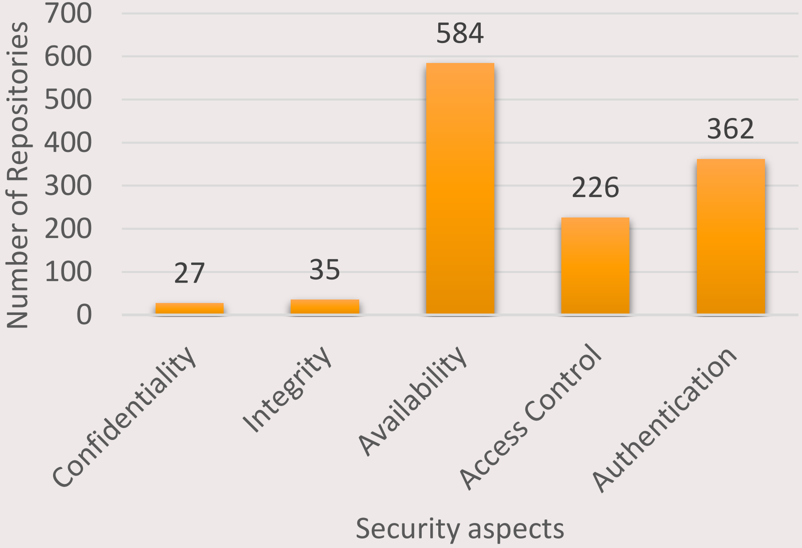
- ➢ The attention to security and attacks in IoT projects is increasing.

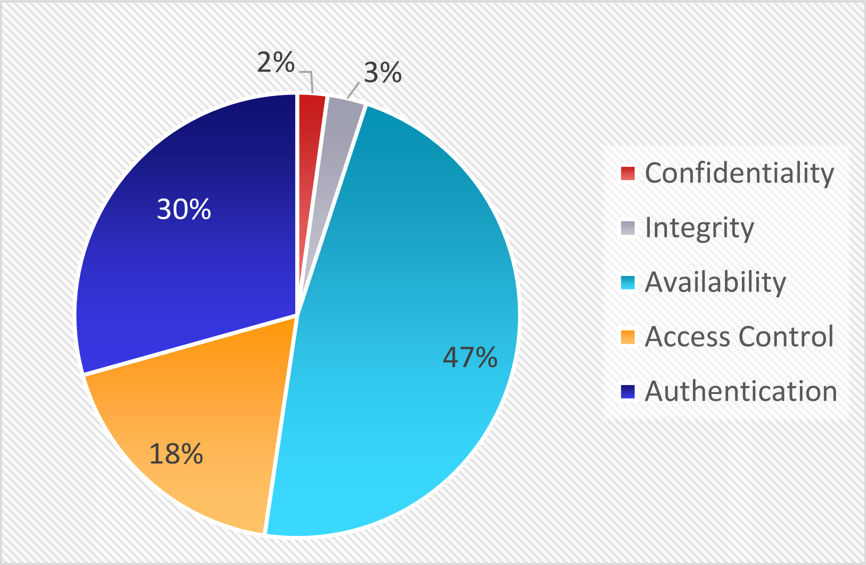
IoT applications





Security aspects among IoT Repos



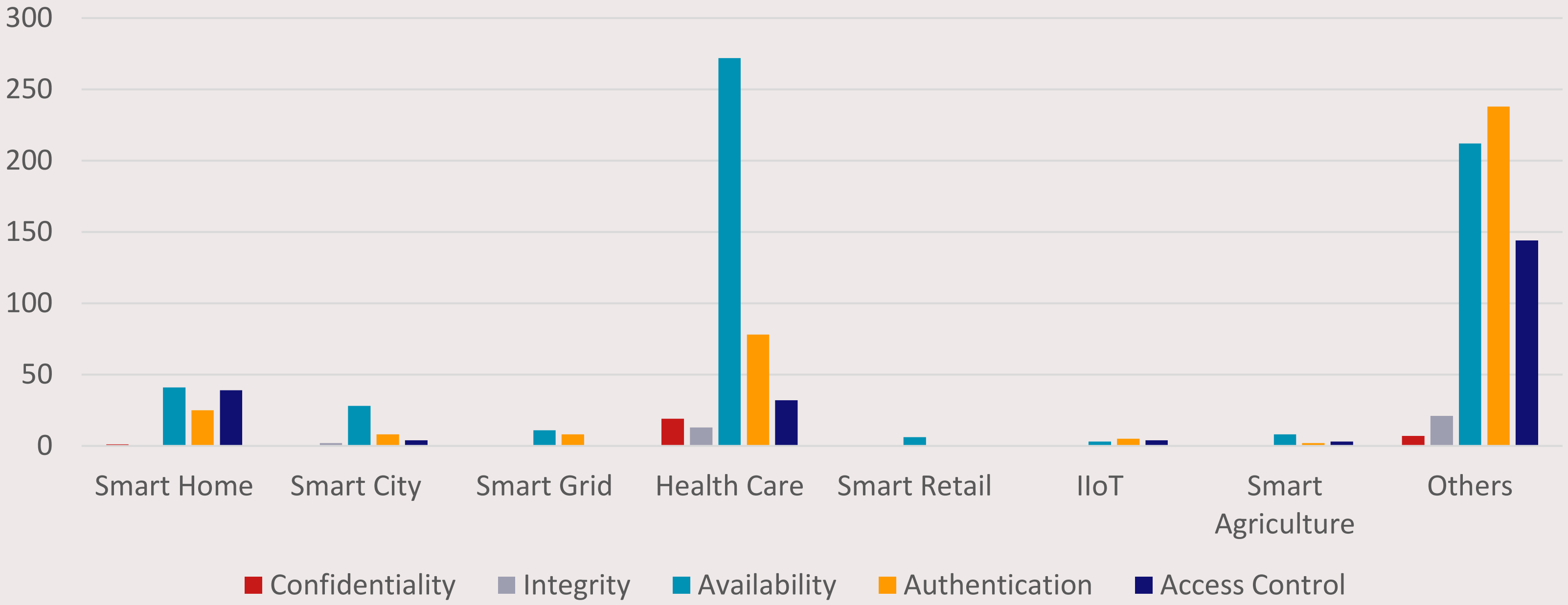


Security aspects among IoT Repos:

Overall analysis

- Emphasis on Availability:
  - The data suggests a **strong emphasis on availability-related concerns**, as evidenced by the significantly higher count for Availability (584 instances) compared to other security properties.
  - The high count for Availability underscores the critical importance of ensuring uninterrupted operation and service delivery in IoT systems.
- Focus on Authentication and Access Control:
  - Authentication (362 instances) and Access Control (226 instances) are **also noticeable concerns**, indicating a focus **on controlling access to IoT devices and resources**.
- Concerns about Confidentiality and Integrity:
  - While Confidentiality (27 instances) and Integrity (35 instances) are mentioned, they appear to be relatively **lower in count compared to Availability, Authentication, and Access Control.**
  - The lower counts for Confidentiality and Integrity may suggest that while important, they may not be perceived as immediate or prevalent concerns in the context of the analyzed data

Security Aspects per IoT application



Security aspects among IoT Repos:

Overall analysis

- Confidentiality:
  - Health Care has the highest emphasis on confidentiality, with 19 instances, indicating the importance of **protecting sensitive patient information.**
  - Other domains also show concerns for confidentiality but to a lesser extent.

- Integrity:
  - Health Care and Smart City domains exhibit the highest emphasis on integrity, with 13 and 2 instances, respectively. This underscores the importance of **ensuring data integrity in critical systems** like **healthcare and urban infrastructure**.

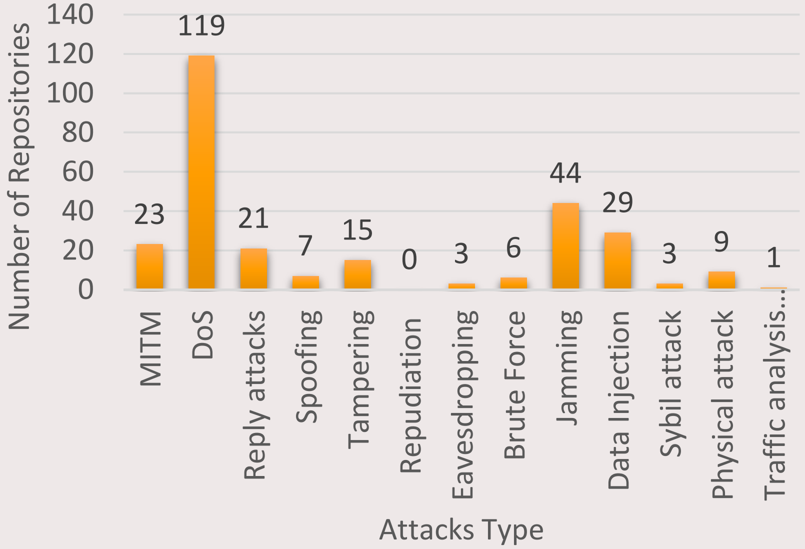
- Availability:
  - Availability is a significant concern across all domains, with Health Care having the highest emphasis, followed by Smart City and Others. This **indicates the criticality of continuous availability in IoT systems**, especially in **healthcare and urban environments.**
  - Even domains like Smart Grid and Smart Agriculture, which may not seem as reliant on real-time availability, still demonstrate a notable emphasis on this aspect.

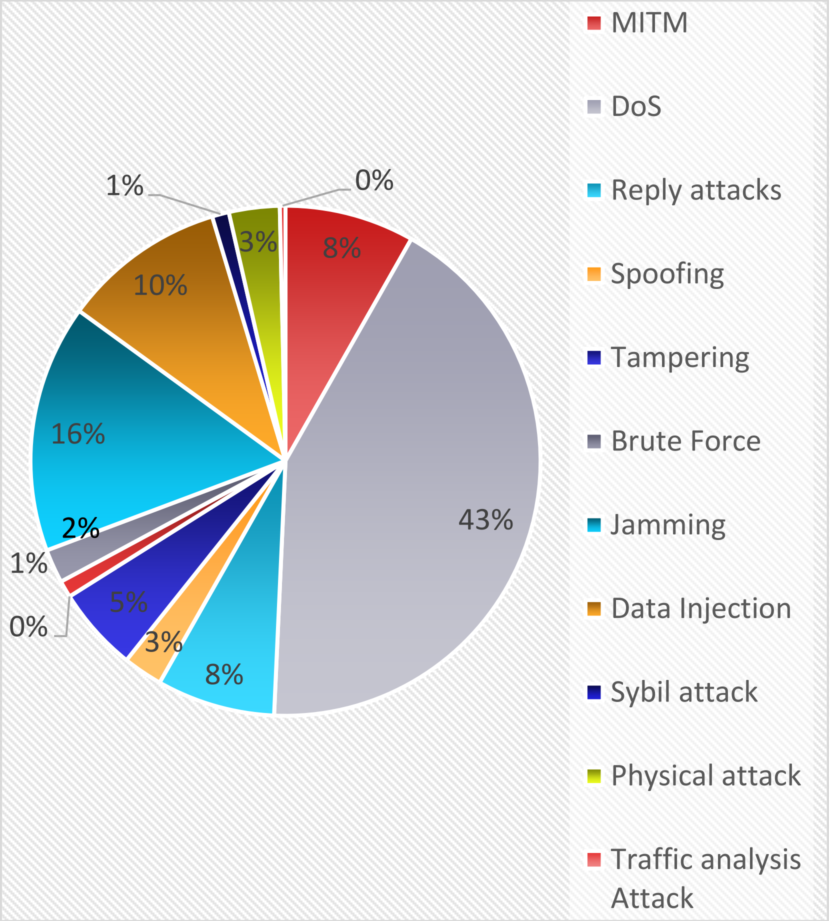
Security aspects among IoT Repos:

Overall analysis

- Authentication:
  - Authentication is crucial across all domains, with healthcare domains showing the highest emphasis. This suggests a strong focus on **ensuring secure access** to IoT systems and services, particularly in healthcare settings.
- Access Control:
  - Access control is a significant concern across in most domains, with Smart home, Health Care and Others demonstrating the highest emphasis. This highlights the importance of **enforcing proper access controls to protect IoT systems and data from unauthorized access.**
  - Domains like Smart Retail and IIoT exhibit minimal emphasis.
- Overall, Health Care emerges as a domain with heightened concerns for confidentiality, integrity, availability, authentication, and access control, reflecting the critical nature of healthcare IoT systems. Other domains also demonstrate varying degrees of emphasis on these security aspects, reflecting the diverse security challenges inherent in IoT deployments.

Attacks among IoT Repos



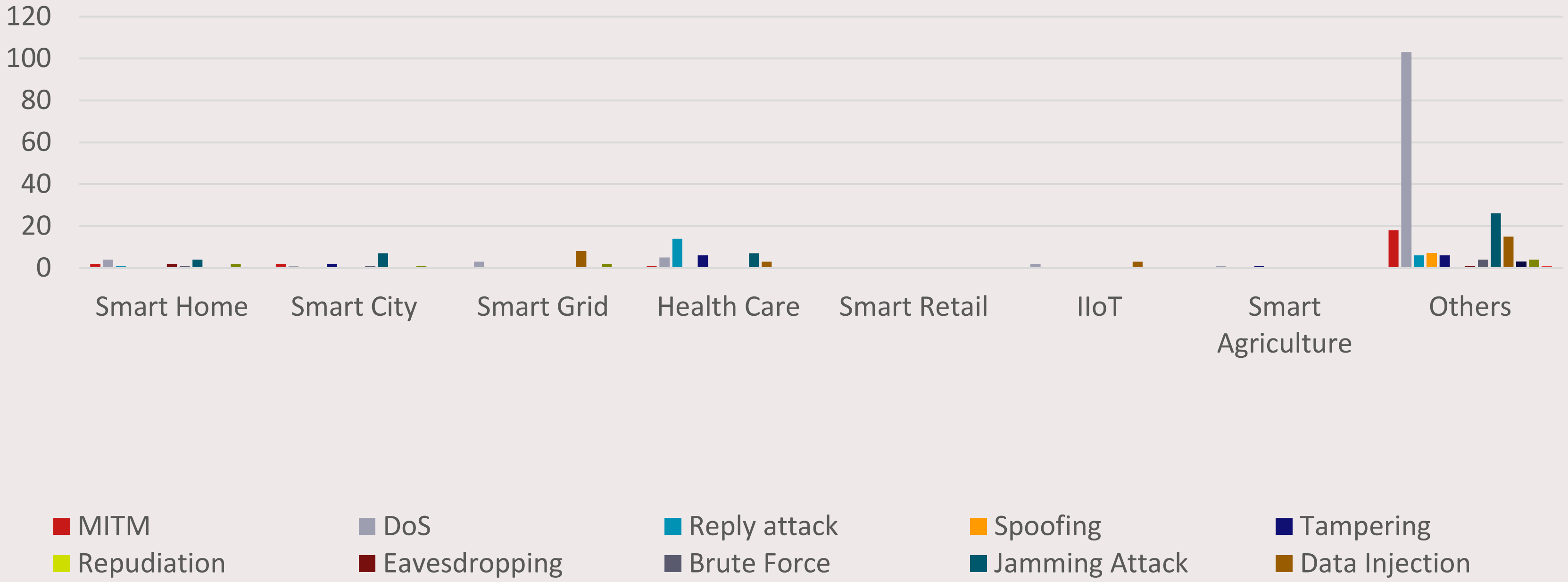


Attacks among IoT Repos: Overall analysis

- Prevalence of various attack types:
  - Denial of Service (DoS) attacks are the most frequent, with 119 instances, showing an **important risk to the availability of IoT services**.
  - Man-in-the-Middle (MITM) attacks are also almost common, with 23 reported cases, indicating **communication channel weaknesses.**

- Significance of Network-Based Attacks:
  - Network-based attacks, such as MITM, DoS, and jamming, are a significant component of the total threat environment.
  - MITM attacks can intercept and modify data sent between IoT devices, compromising data confidentiality and integrity.
  - DoS attacks try to interrupt IoT services by flooding networks or systems with excessive traffic.
  - Jamming attacks target communication channels, which can disrupt the delivery of important information in IoT systems.

Attacks per IoT application



Attacks among IoT Repos: Overall analysis

- Domain-specific Vulnerabilities:
  - Each domain exhibits varying levels and types of attacks.
    - For example, **Healthcare** has higher instances of Reply Attack, Tampering, and Repudiation, indicating vulnerabilities related to data **integrity, privacy, and authentication**.
    - **Smart City** faces challenges with MITM attacks, Tampering, and Data Injection, which could compromise **the integrity and reliability of city infrastructure and services**.
- Common Security Threats:
  - Some attack types appear **across multiple domains**.
    - For instance, MITM attacks are observed in Smart Home, Smart City, and IIoT, suggesting a common vulnerability in communication protocols or network architectures.
  - DoS attacks are prevalent in Healthcare and Smart Grid, indicating risks to service availability and operational continuity in critical infrastructure.
- Critical Infrastructure Risks:
  - Certain domains, like Healthcare and Smart City, experience attacks that can **have significant implications for public health and safety**. For example, attacks targeting Healthcare systems could compromise **patient data privacy and disrupt medical services**.

Attacks among IoT Repos: Overall analysis

- MITM (Man-in-the-Middle):
  - The highest occurrence is observed in the **"Smart Heath"** domain, indicating **potential vulnerabilities in health care systems**.
  - The "Smart home" domain also experiences MITM attacks, suggesting risks to **in home automation systems and communication networks.**
- DoS (Denial of Service):
  - The **"Health Care"** domain shows the highest frequency, indicating **potential disruptions in healthcare services and patient care delivery**.
  - Additionally, the "Smart Grid" domain experiences significant occurrences, highlighting risks to critical infrastructure stability.
- Reply Attack:
  - Most prevalent in the **"Health Care"** domain, suggesting **vulnerabilities in medical IoT systems and potential compromises to patient data integrity.**
  - Relatively low occurrences in other domains, indicating a specific concern within healthcare IoT systems.

Attacks among IoT Repos: Overall analysis

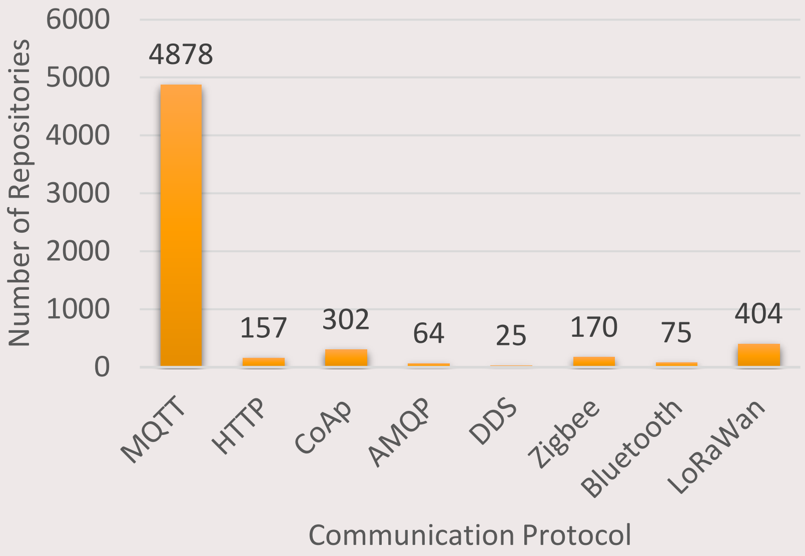
- Spoofing:
  - Predominantly found in the **"Health Care"** domain, posing risks to patient identification and medical device authentication.
  - Also present in the "IIoT" domain, indicating potential threats to industrial control systems and manufacturing processes.

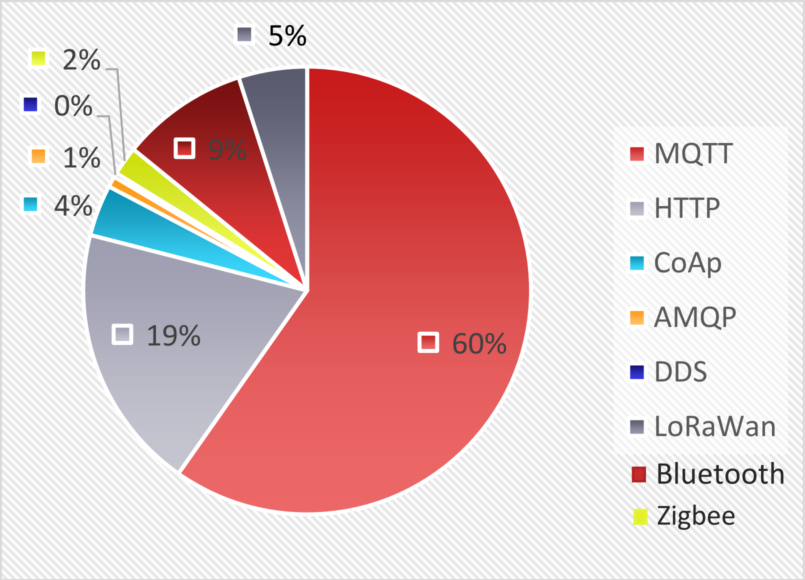
- Tampering:
  - The "Health Care" domain exhibits the highest occurrences, where tampering attacks can manipulate medical records or device functionality, jeopardizing patient safety.
  - Also observed in the "Smart City" domain, suggesting risks to infrastructure integrity and public safety.

Attacks among IoT Repos: Overall analysis

- Repudiation:
  - Low occurrences across most domains, except for "Others," where it has moderate frequency. Repudiation attacks may pose challenges in establishing accountability and non-repudiation mechanisms.
- Eavesdropping:
  - Occurs in "Smart Home" and "Others" but generally low across all domains. Eavesdropping attacks can compromise privacy and sensitive data confidentiality in IoT communication channels.
- Brute Force:
  - Moderate occurrences in "Smart Home," "Smart City," and "Others," indicating potential vulnerabilities in access control mechanisms and password security across various IoT systems.
- Jamming Attack:
  - The "Smart Grid" domain experiences significant occurrences, where jamming attacks can disrupt power grid communication and control systems, leading to service interruptions and potential cascading failures.

Communication Protocol among IoT Repos





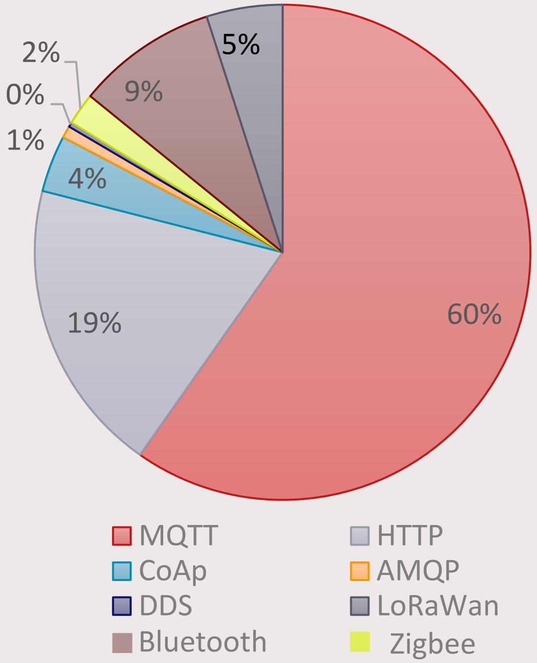
Communication Protocol among IoT Repos：

Overall analysis

- MQTT:
  - MQTT is the most widely used communication protocol among the Repos. It's known for its lightweight nature and efficiency in IoT applications.
  - From a security perspective, MQTT supports features like **TLS encryption for secure communication and authentication** mechanisms, which are essential for **protecting IoT devices and data**.

- HTTP:
  - Although HTTP is widely used on the web, its usage in IoT applications is relatively low compared to MQTT.
  - From a security standpoint, HTTP can be secured using HTTPS (HTTP Secure), which provides **encryption through SSL/TLS**, but it **may not be suitable for resource-constrained IoT devices due to overhead.**

- CoAP:
  - CoAP is designed specifically for IoT devices with limited resources. It's lightweight and supports RESTful principles, making it suitable for constrained environments.
  - From a security perspective, CoAP can **utilize Datagram Transport Layer Security (DTLS) for encryption and authentication**, providing **secure communication for IoT devices**.



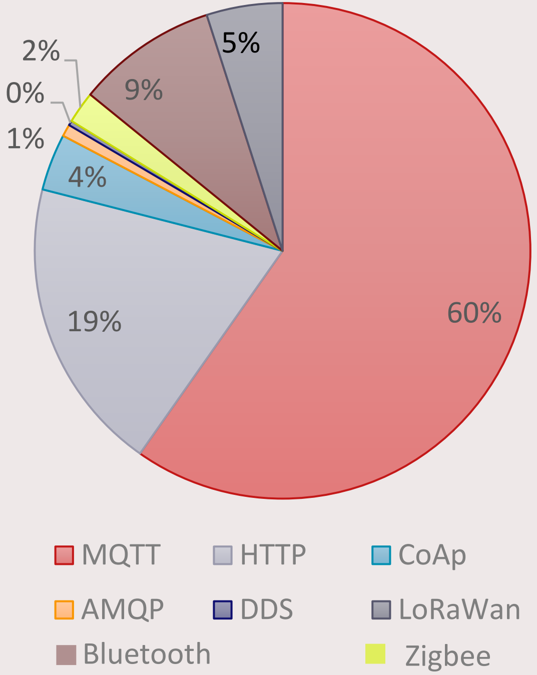
Communication Protocol among IoT Repos：

Overall analysis

- AMQP:
  - AMQP is a messaging protocol that supports efficient message queuing and routing. It's less common in IoT applications compared to MQTT and CoAP.
  - From a security standpoint, AMQP can leverage **TLS for encryption and authentication, ensuring secure communication channels between devices and brokers.**

- DDS (Data Distribution Service):
  - DDS is a messaging standard often used in real-time and mission-critical applications. While less prevalent in IoT compared to other protocols, it offers features like Quality of Service (QoS) and reliability.
  - Security in DDS implementations typically **involves access control and data encryption to protect sensitive information**.

- Zigbee:
  - Zigbee is a wireless communication protocol commonly used in home automation and industrial applications. It operates on low-power devices and supports mesh networking.
  - Security in Zigbee networks **involves encryption, authentication, and secure key management to prevent unauthorized access and ensure data confidentiality**.

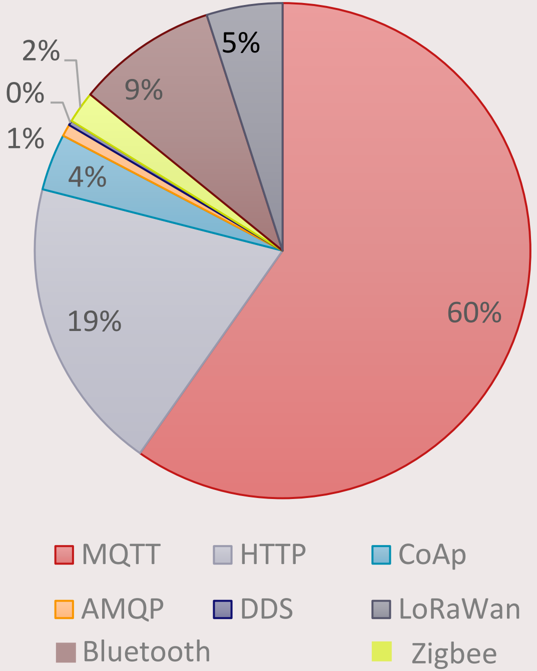


Communication Protocol among IoT Repos：

Overall analysis

- Bluetooth:
  - Bluetooth is widely used for short-range wireless communication in IoT devices such as wearables, smart home devices, and beacons.
  - Bluetooth offers various security features such as **pairing mechanisms, encryption, and authentication to protect communication between devices.**

- LoRaWAN:
  - LoRaWAN is a low-power, wide-area networking protocol designed for long-range communication in IoT applications. It's commonly used in applications like smart cities and agricultural monitoring.
  - LoRaWAN incorporates security features such as **end-to-end encryption and device authentication to ensure secure communication over long distances.**



Conclusion

Conclusion

- This study presents a landscape of IoT software projects publicly available on GitHub.
- In this research, ≈296.978K IoT repositories were identified and categorized.
- The security properties, common attacks in IoT systems, and common communication protocols were analyzed.
- Overall, the analysis underscores the importance of understanding the security landscape of IoT across different application domains and implementing targeted security strategies to mitigate risks and safeguard critical infrastructure and services.
- This research is the first of its kind and pays attention to IoT systems, attacks, and security aspects of them.

Next Steps

- Picking up some examples and know what kind of solutions this repos provide against certain attacks.
- Go deep further to understand
  - How software that makes secure IoT is developed
  - What kind of solutions they provided
- How we can help developing this software more secure

Thank you!