DETECTING ANOMALY-BASED CYBERATTACKS IN SMART HOMES: A COMPREHENSIVE REVIEW OF THE LITERATURE

Dr. Abu Salim K

ABSTRACT

The paper explores the vulnerability of smart homes to cyberattacks due to the collection of sensitive data, emphasizing the need for effective anomaly detection. While existing literature often addresses IoT-related cyber threats, there is a notable gap in focusing on anomalies specific to smart homes. The study conducts a systematic literature review, offering an adapted taxonomy for classifying anomaly detection methods. The findings reveal a growing interest in utilizing anomaly-based models, particularly centralized and network-based features, for detecting cyberattacks in smart homes. Popular techniques include ensemble and deep learning methods. However, challenges such as limited diversity in existing datasets and the absence of comprehensive datasets representing smart home complexity underscore the need for further research to enhance detection model generalizability.

Index Terms *Anomaly detection, Machine Learning, Internet of Things (IoT), Smart home, Cyber-security, Cyber-attacks, Systematic literature review (SLR).*

Reference *to this paper should be made as follows**:** Abu Salim K, (2023), "Detecting Anomaly-Based Cyberattacks in Smart Homes: A Comprehensive Review of the Literature" Int. J. Electronics Engineering and Applications, Vol. 11, Issue I, pp. 31-41.*

Biographical notes:

Abu Salim *received his B.Tech (CSE) from Dr. RML Avadh University, Faizabad in 2003, M.Tech from JNTU Kakinada in 2009 and PhD in cloud computing from Glocal University , Saharanpur, UP in 2023. Currently he is working as Lecturer in the Jazan University, KSA. He has published 12 papers in journals and conference proceedings. His research is focused on cloud computing, data security and, Machine Learning.*

1. INTRODUCTION

The accelerating digitization of society has brought forth a concomitant increase in concerns regarding the safeguarding of digital assets from an ever-expanding array of cyber threats. These threats encompass a wide range of valuable data, including operational data, Personally Identifiable Information (PII), and strategically sensitive information. The growing prevalence of telecommuting further intensifies the need for securing smart homes, as these environments, often characterized by user inexperience and the integration of insecure Internet of Things (IoT) devices, become particularly appealing targets for cybercriminals.

In response to the evolving cyber threat landscape, Anomaly-Based Detection Systems (ADS) have emerged as a promising strategy. These systems operate by distinguishing abnormal patterns of behavior from normal patterns, allowing for the identification of novel cyberattacks. While ADS has demonstrated success in the realm of IoT security, applying this approach to the unique context of smart homes presents notable challenges. The challenges stem from the diverse nature of user-device interactions and the dynamic, ever-changing environment within smart homes.

A systematic literature review (SLR) conducted in this study makes a substantial contribution to the field by providing a comprehensive exploration of cybersecurity threats in smart homes. The review goes beyond threat identification to include an in-depth analysis of Anomaly-Based Detection Systems tailored for smart home environments. It introduces a taxonomy that classifies various anomaly detection methods, shedding light on the diverse approaches employed to secure these complex environments.

The SLR also extends its scope to encompass an examination of associated issues, challenges, and gaps in current research and practices. By identifying these challenges, such as the intricate interplay between users and devices and the dynamic nature of smart home environments, the review lays the foundation for understanding the complexities that must be addressed in the pursuit of effective cybersecurity measures.

The article, structured methodically, progresses from a detailed exploration of cybersecurity threats in smart homes to an in-depth analysis of Anomaly-Based Detection Systems, providing readers with a nuanced understanding of the current state of the field. The taxonomy introduced for classifying anomaly detection methods facilitates a structured overview, aiding researchers, practitioners, and policymakers in navigating the landscape of smart home cybersecurity.

The review concludes by offering insights gleaned from the findings and proposes research directions for the future. These proposed directions aim to fill existing gaps in understanding and address the challenges posed by securing smart homes effectively. By doing so, the study not only serves as a valuable resource for current cybersecurity practitioners and researchers but also lays the groundwork for advancing the field, ensuring the security of the increasingly interconnected and digitized smart home environments in the years to come.

2. BACKGROUND

The systematic literature review (SLR) outlined in the article sets out to comprehensively explore the intersection of smart homes, cyber threats, and anomaly detection. Smart homes are defined by their integration of intelligent devices and services, creating environments that aim to enhance efficiency and convenience for residents. The study recognizes the pivotal role that these interconnected systems play in modern living, from automated lighting and security systems to smart appliances and entertainment devices.

To provide a structured understanding of the cyber threats faced by smart homes, the SLR introduces a taxonomy based on the STRIDE model. This model categorizes threats into different types, including Denial of Service (DoS), flooding, probing, and more. The taxonomy serves as a valuable tool for classifying and organizing the diverse range of cyber threats that smart homes encounter, offering a systematic approach to analyzing potential risks.

The article delves into the intricate landscape of cyberattacks on smart homes, encompassing both traditional threats, such as brute force attacks, and modern threats like ransomware and attacks specifically targeting Internet of Things (IoT) devices. The diverse nature of these cyber threats highlights the need for a multifaceted security approach that can address the evolving tactics employed by malicious actors.

In the realm of Intrusion Detection Systems (IDS), the article provides a comprehensive exploration. It discusses various detection strategies and placement strategies, considering whether IDS should be centralized, distributed, or follow a hybrid approach. This discussion underscores the importance of strategically deploying IDS to effectively monitor and identify potential threats within smart home environments. Validation strategies are also considered, emphasizing the need to ensure the accuracy and reliability of intrusion detection mechanisms.

Anomaly detection emerges as a key focus within the SLR, with the article introducing different approaches categorized as supervised, semi-supervised, and unsupervised. The taxonomy for anomaly detection methods further breaks down these approaches into classification-based, clustering-based, information theory-based, and other methods. This categorization provides a structured overview of the various techniques employed to detect abnormal behavior within smart home systems, allowing for a nuanced understanding of the methodologies used to enhance security.

The overarching goal of the SLR is to offer insights into the landscape of cyber threats faced by smart homes, the methodologies employed for anomaly detection, and the inherent challenges in securing these environments. By providing a structured and comprehensive overview, the article serves as a valuable resource for researchers, practitioners, and policymakers seeking to understand and address the complexities of cybersecurity in the context of smart homes. Additionally, the study lays the groundwork for future research directions, paving the way for advancements in securing smart home environments against an ever-evolving array of cyber threats.

3. THE REVIEW PROCESS

The article employs a rigorous Systematic Literature Review (SLR) methodology to present a comprehensive overview of the state of the art in Anomaly Detection Systems (ADS) within smart home environments. The SLR methodology is structured into distinct phases, each designed to ensure methodological rigor, objectivity, and transparency throughout the review process.

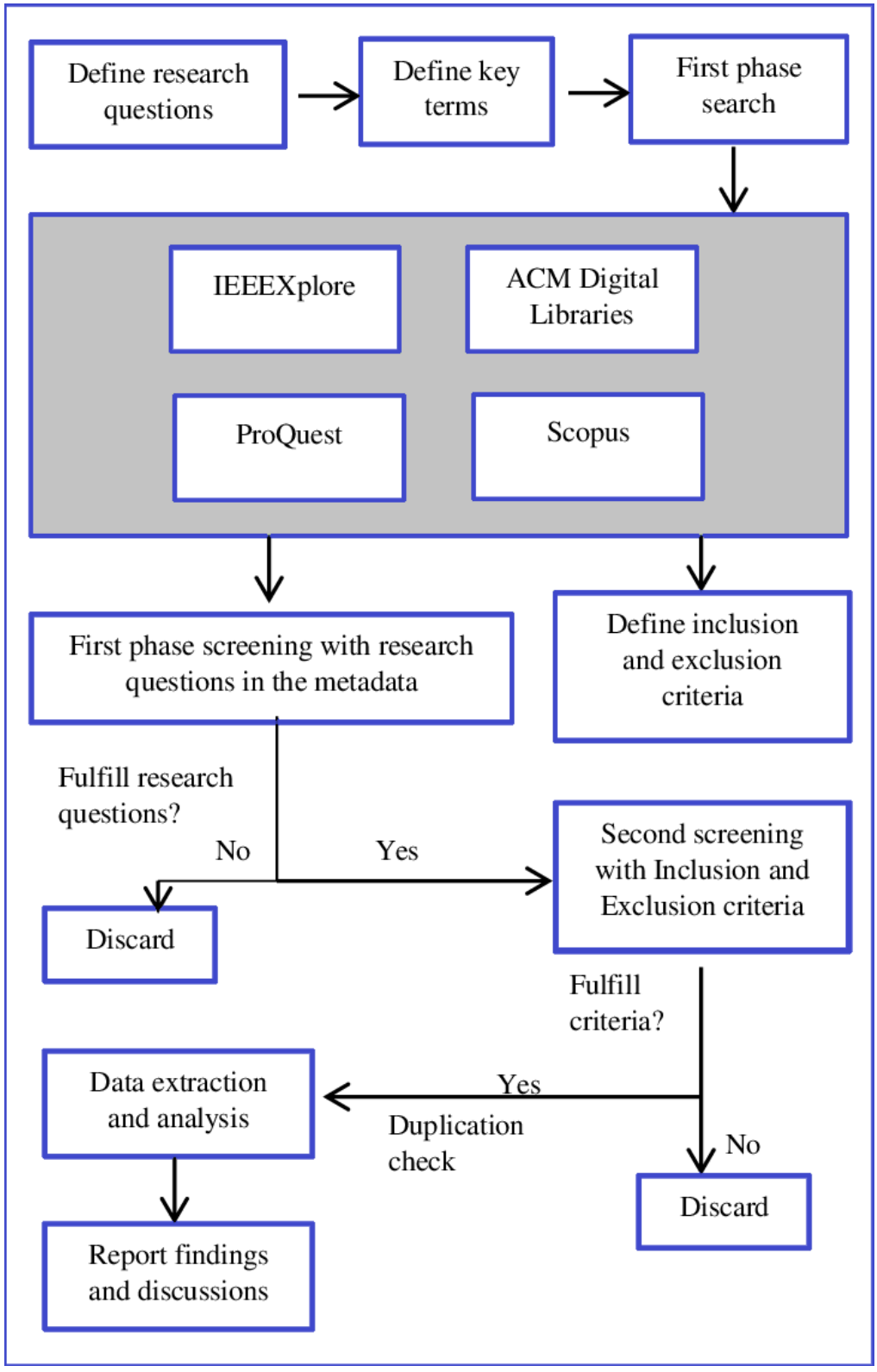

Figure 2: Phases of a Systematic Literature Review.

The planning phase serves as the foundation, where the need for the review is identified, and research questions are formulated. Research questions (RQs) are carefully crafted to cover essential aspects of ADS in smart homes, including the volume of research, strategies employed, identification of best methods, evaluation metrics, datasets used, and future challenges. Defining a review protocol is crucial at this stage, establishing guidelines and criteria to be adhered to during the review process.

Moving to the conducting phase, a comprehensive search strategy is outlined, incorporating relevant keywords and leveraging prominent digital libraries such as Web of Science, Scopus, IEEE Xplore, and ACM Digital Library. The article selection process involves meticulous steps such as eliminating non-relevant papers, removing duplicates, and applying predefined inclusion and exclusion criteria. Quality assessment rules are established, and a set of questions is employed to systematically evaluate the quality of the selected articles, ensuring that only high-quality and relevant contributions are included in the review.

The data extraction process is a critical component of the review, involving the systematic gathering of information from selected papers. Key information includes the method with the best results, detection strategies, placement and validation strategies, source of features, datasets used, anomaly data sources, and evaluation metrics. This meticulous data extraction process forms the basis for synthesizing the gathered information to address each research question systematically.

The results of the conducting phase demonstrate the effectiveness of the review protocol, as the number of papers is significantly reduced from an initial pool of 1438 to a refined set of 98 articles. The selected articles are systematically listed, providing transparency and allowing readers to trace the inclusion process. The data synthesis process involves summarizing the gathered information in a structured manner, addressing each research question comprehensively and providing a cohesive overview of ADS in smart home environments.

4. RESULTS

4.1. RQ1: Volume of Research

The volume of research on anomaly detection of cyberattacks in smart home contexts has experienced a notable increase, especially during the period from 2019 to 2022. This upsurge suggests a growing interest and emphasis on securing smart home environments. The heightened attention during this period may be attributed to the increased focus on telecommuting induced by the COVID-19 pandemic, highlighting the relevance and urgency of securing home networks. Conferences, particularly those included in the IEEE Xplore Database, have emerged as the primary platforms for disseminating research findings, underlining the significance of academic and industry collaboration in advancing knowledge in this domain.

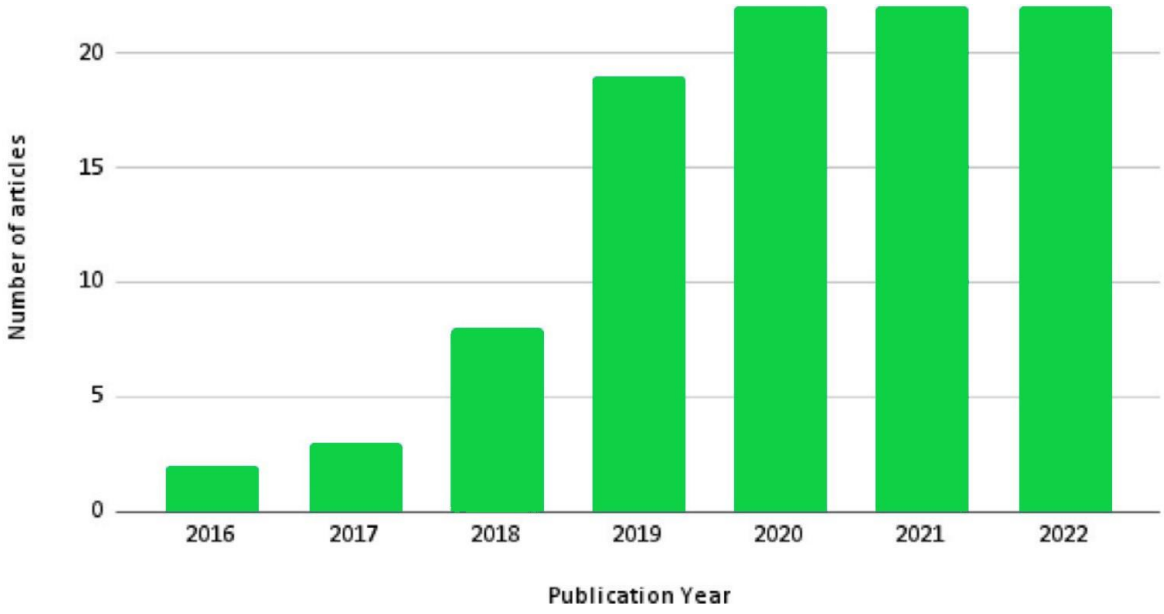

Figure 1: Distribution of Publications by Year.

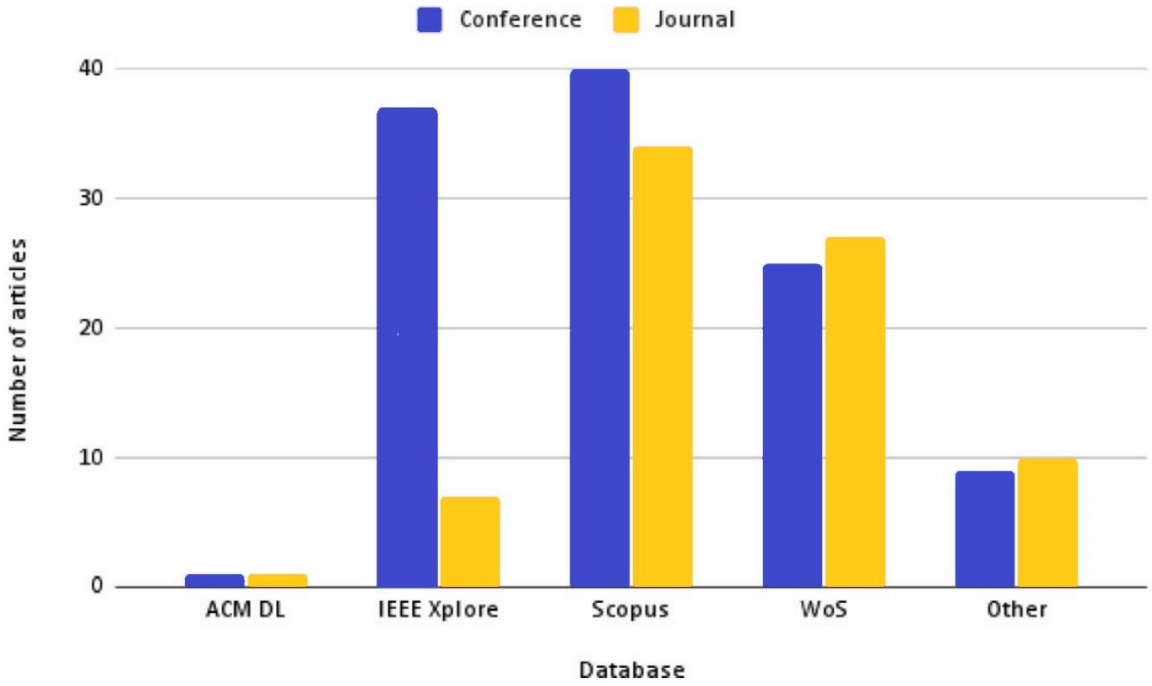

Figure 2: Distribution of Published Papers.

4.2. RQ2: Detection, Placement, and Validation Strategies

Network-based detection, centralized placement, and empirical validation strategies dominate the landscape of anomaly detection systems (ADS) in smart home environments. Approximately 74.5% of the reviewed papers leverage network-based detection methodologies, emphasizing the importance of monitoring network activities for potential threats. The centralized approach is favored by around 90% of researchers, highlighting its practicality, especially in single-home scenarios. However, the discussion recognizes the potential suitability of distributed approaches, particularly for securing multiple homes, although considerations of processing capacity become paramount.

4.3. RQ3: Best ADS Methods

Ensemble learning, with a particular emphasis on random forest, and deep learning methods, such as convolutional neural networks (CNN) and autoencoders, emerge as the best-performing and widely adopted methods for anomaly detection in smart homes. Ensemble learning and deep learning have gained prominence in recent years, showcasing their efficacy in adapting systems to changing environments. Techniques like reinforcement learning and transfer learning are highlighted, underlining the community's commitment to advancing methods capable of evolving with the dynamic nature of smart home environments.

4.4. RQ4: Evaluation Metrics

The evaluation of anomaly detection systems in smart homes commonly employs metrics such as precision, accuracy, recall, f-score, and detection rate. Beyond traditional metrics, there is a growing focus on time and resource usage metrics, acknowledging the constraints and limitations in the resource-scarce smart home environments. Context-aware metrics, including f-score, have gained relevance as researchers increasingly consider the unique characteristics of IoT environments in their assessments.

4.5. RQ5: Source Datasets and Features

The selection of datasets for evaluating anomaly detection systems in smart homes varies, with nearly half of the researchers relying on self-generated datasets. This underscores the recognition of the need for problem-specific datasets to accurately reflect the intricacies of real smart home environments. Public datasets, such as IoT Botnet, IoT Network, and N-BaIoT, are commonly utilized, demonstrating a balance between the creation of tailored datasets and leveraging established datasets. The majority of datasets focus on network traffic-derived characteristics due to their ease of acquisition and applicability across diverse smart home setups.

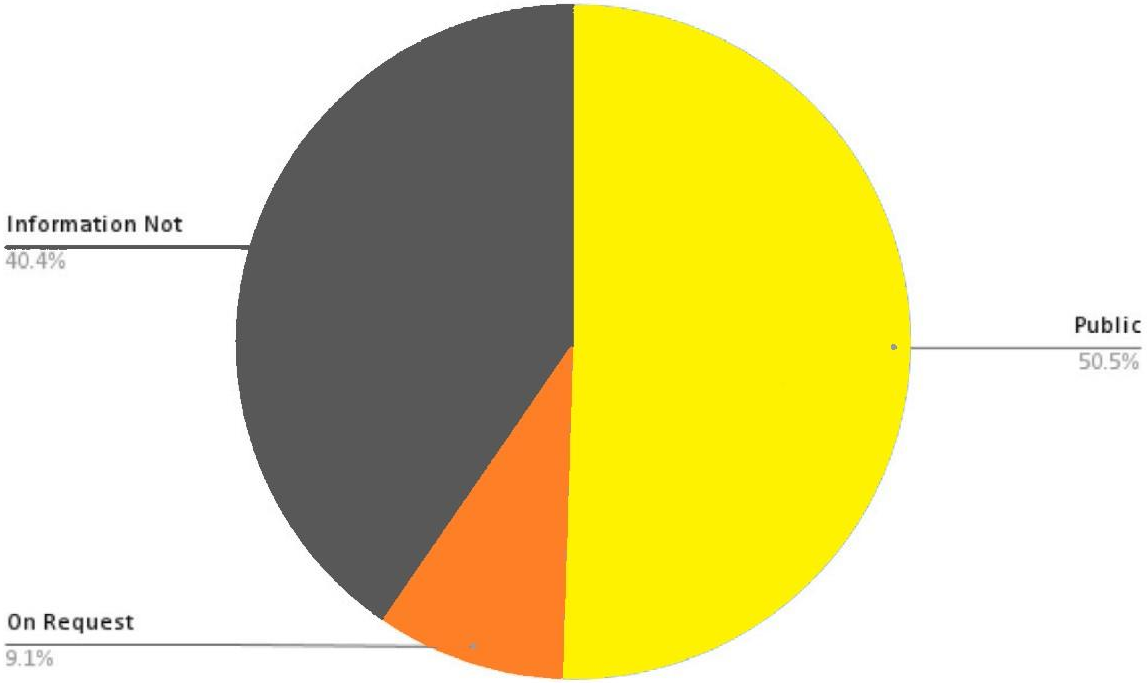

Figure 3: Availability of Datasets.

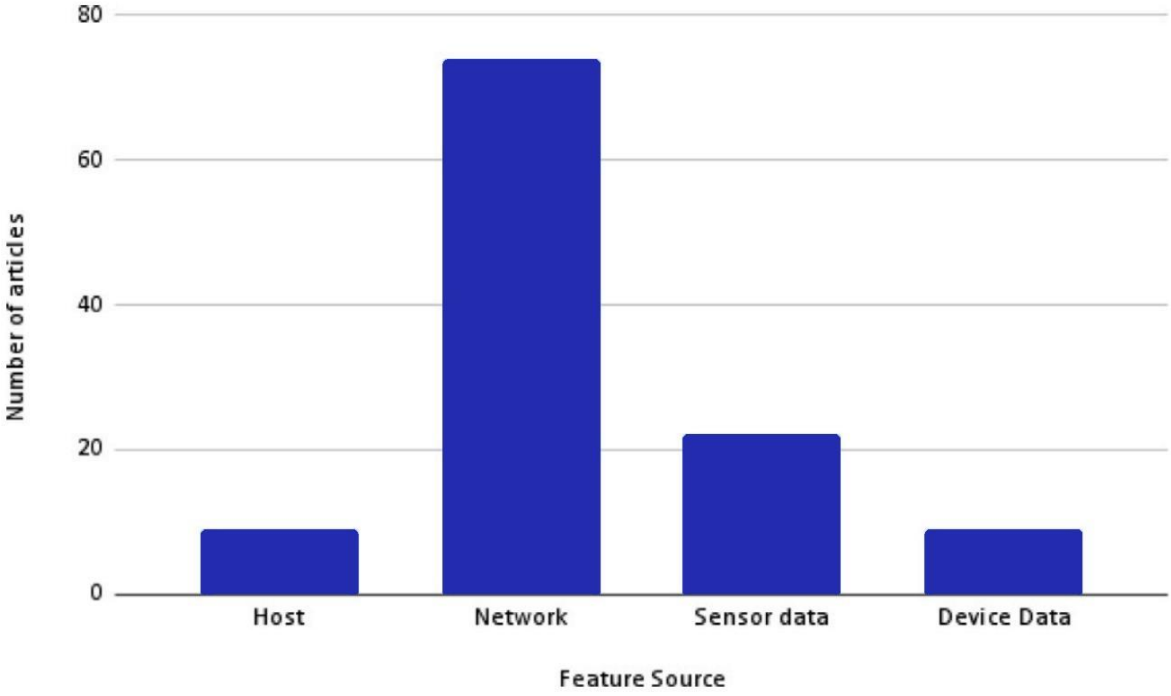

Figure 4: Features of Datasets.

4.6. RQ6: Research Gaps and Future Challenges

Identified challenges include the necessity for diverse datasets that truly mirror real smart home environments, deployment challenges in real-world settings, profiling normal device behavior, recognizing a variety of attacks (including unknown ones), and the integration of additional machine learning models to enhance detection capabilities. Future challenges involve the evaluation of transfer techniques, adapting models to Software-Defined-Network (SDN) environments, exploring lightweight procedures, anonymizing data for confidentiality, incorporating additional features, and building adaptive models. Researchers express interest in combining multiple methods to improve detection rates and stress the importance of evaluating models on additional datasets to ensure generalizability and robustness. These challenges and future directions collectively contribute to shaping the ongoing discourse and evolution of anomaly detection in smart home environments.

5. DISCUSSION

5.1. Architecture:

Developing an effective Intrusion Detection System (IDS) architecture for smart homes involves addressing various challenges to ensure adaptability, user intervention avoidance, multi-user support, scalability, real-time operation, lightweight design, and high performance against both known and zero-day cyberattacks. Smart homes, characterized by a dynamic environment with diverse user behaviors, require IDS architectures capable of continuous adaptation. The architecture should integrate classification models that can evolve based on emerging insights into evolving attack methods and changes in user behaviors. Centralized devices with enhanced resources, leveraging fog computing, present a potential solution to consolidate data and improve sensing performance, contributing to the scalability and real-time operation of the IDS.

5.2. Classification Methods:

The deployment of Deep Learning (DL) methods in smart homes demands caution due to the presence of devices with limited processing power. The suitability of fog computing as a means to alleviate constraints on devices must be carefully evaluated, taking into consideration potential oversights in detecting attacks occurring inside networks. Balancing the need for sophisticated classification methods with the constraints of resource-limited devices is crucial for the effectiveness of intrusion detection in smart home environments.

5.3. Smart Home Features:

The modeling of smart homes often relies on characteristics derived from network traffic, with limited attention given to context-specific features. Identifying relevant features that can accurately model both normal and malicious behavior in smart homes remains a challenge. A comprehensive understanding of smart home features is essential for the development of accurate and reliable anomaly detection models that can distinguish between legitimate and malicious activities in diverse scenarios.

5.4. Privacy:

Privacy considerations are paramount in the development of anomaly detection techniques that rely on large datasets potentially containing sensitive information. While many articles acknowledge the relevance of privacy, few actively implement measures to protect it. The integration of privacy protection measures, including encryption, anonymization, pseudo-anonymization, and privacy by

design, is essential in IDS development for smart homes. Striking a balance between effective intrusion detection and the protection of user privacy is critical for the widespread acceptance and ethical deployment of such systems.

5.5. Smart Home Datasets:

Existing datasets used for training and evaluating anomaly detection systems often lack diversity in terms of cyberattacks and the interaction between people and devices. To advance the field, there is a pressing need for publicly accessible datasets sourced from real smart homes. These datasets should not only contain a variety of cyberattacks but also reflect the changing behaviors of inhabitants over extended periods. The development of diverse and representative datasets is fundamental for training anomaly detection models that can effectively generalize to real-world smart home scenarios.

5.6. Changing Environment:

Smart home environments are inherently dynamic, experiencing continuous changes in device configurations, user behaviors, and network interactions. Anomaly Detection Systems (ADS) for smart homes need to be adaptive and resilient to handle this dynamic nature effectively. Incremental learning, unsupervised learning, transfer learning, and reinforcement learning are proposed as strategies to develop ADS that can adapt to evolving environments, ensuring consistent and reliable intrusion detection.

5.7. Zero-Day Cyberattacks:

Anomaly-based detection is effective against unknown attacks, commonly referred to as zero-day attacks. However, few articles thoroughly evaluate the models' ability to classify zero-day attacks as anomalies. It is essential to broaden the evaluation methodologies to encompass a wider spectrum of potential anomalies, including those that may not necessarily be direct cyberattacks. This includes considerations for anomalies arising from device malfunctions or changes in user behavior, ensuring that the ADS can distinguish between legitimate deviations and actual security threats. Evaluating models against a diverse set of scenarios and categories is crucial to achieving comprehensive coverage in identifying zero-day cyberattacks and preventing false positives.

6. CONCLUSION

In recent years, there has been a notable surge in research focused on developing anomaly-based Cyberattack Detection Systems (ADS) for smart homes. This trend is propelled by the increasing prevalence of telecommuting, which has accentuated the significance of detecting both known and unknown cyber threats in the context of intelligent home environments. This paper provides a comprehensive review of the recent advancements in ADS research, offering insights into various aspects crucial for enhancing the cybersecurity posture of smart homes.

The review covers a spectrum of key elements within the domain of anomaly-based cyberattack detection in smart homes. It delves into detection strategies, emphasizing the need for systems capable of identifying both familiar and novel threats. The research recognizes that the unique features of smart homes, such as multiple inhabitants, evolving behaviors, intricate device interactions, cloud-based device management, and a user base often unaware of security measures, necessitate tailored intrusion detection systems.

Placement and validation methods are explored, addressing the challenge of optimally situating detection mechanisms within the smart home architecture. Recognizing the dynamic nature of smart homes, where user behaviors and device interactions continually evolve, the paper underscores the importance of holistic intrusion detection system architectures. Such architectures should not only focus on identifying recent attacks but also consider user-centric concerns, acknowledging the diverse and evolving nature of smart home environments.

Data sources and datasets play a crucial role in training and evaluating ADS. The review acknowledges the need for tailored datasets that capture the intricacies of smart home interactions, including diverse cyber threats and the evolving behaviors of inhabitants over extended periods. Evaluation metrics are discussed, with an emphasis on metrics that reflect the limited resources in smart home contexts, ensuring a balanced assessment of system performance.

The study also sheds light on emerging challenges, recognizing that the landscape of smart home cybersecurity is continually evolving. The holistic approach advocated in the paper encompasses considerations beyond the technical aspects of attack detection, extending to user-centric concerns and the need for comprehensive IDS architectures.

REFERENCES

[1] F. Dahlqvist, M. Patel, A. Rajko, J. Shulman, Growing opportunities in the Internet of Things, McKinsey & Company, 2019, pp. 1–6.

[2] R. Singh, H. Kumar, R.K. Singla, R.R. Ketti, Internet attacks and intrusion detection system: A review of the literature, Online Inform. Rev. 41 (2) (2017) 171–184, http://dx.doi.org/10.1108/OIR-12-2015-0394.

[3] M.L. Santos, J.C. Carneiro, A.M. Franco, F.A. Teixeira, M.A. Henriques, L.B. Oliveira, FLAT: Federated lightweight authentication for the Internet of Things, Ad Hoc Netw. 107 (2020) 102253.

[4] OpenID connect basic client implementer's guide 1.0 - draft 42, 2022, https://openid.net/specs/openid-connect-basic-1_0.html.

[5] P. Lade, R. Ghosh, S. Srinivasan, Manufacturing analytics and industrial internet of things, IEEE Intell. Syst. 32 (3) (2017) 74–79, http://dx.doi.org/10. 1109/MIS.2017.49.

[6] C. Keßler, R.T.A. de Groot, Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap, in: D. Vandenbroucke, B. Bucher, J. Crompvoets (Eds.), Geographic Information Science At the Heart of Europe, Springer International Publishing, Cham, 2013, pp. 21–37.

[7] B. Kocacinar, B. Tas, F.P. Akbulut, C. Catal, D. Mishra, A real-time CNN-based lightweight mobile masked face recognition system, IEEE Access 10 (2022) 63496–63507.

[8] E. Manor, S. Greenberg, Custom hardware inference accelerator for TensorFlow lite for microcontrollers, IEEE Access 10 (2022) 73484–73493.

[9] Fazli Wahid, et al., Energy consumption optimization and user comfort maximization in smart buildings using a hybrid of the firefly and genetic algorithms, Energies 13 (17) (2020) 4363.

[10] Naeem Iqbal, Do Hyeun Kim, et al., Iot task management mechanism based on predictive optimization for efficient energy consumption in smart residential buildings, Energy Build. 257 (2022) 111762.

[11] Famous DDoS attacks | Biggest DDoS attacks, Cloudflare, URL https://www.cloudflare.com/learning/ddos/famous-ddos-attacks/.

[12] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J.A. Halderman, L. Invernizzi, M. Kallitsis, Understanding the mirai botnet, in: 26th $USENIX$ Security Symposium ($USENIX$, Security 17, 2017, pp. 1093–1110.

[13] X. Xu, I. Weber, M. Staples, Architecture for Blockchain Applications, Springer International Publishing, Cham, 2019, http://dx.doi.org/10.1007/978-3- 030-03035-3.

[14] T. Domínguez-Bolaño, O. Campos, V. Barral, C.J. Escudero, J.A. García-Naya, An overview of IoT architectures, technologies, and existing open-source projects, Internet Things 20 (2022) 100626, http://dx.doi.org/10.1016/j.iot.2022.100626.

[15] G. Escalas Rodríguez, Diseño Y Desarrollo De Un Prototipo De Riego Automático Controlado Con Raspberry Pi Y Arduino (B.S. thesis), Universitat Politècnica de Catalunya, 2015.

[16] J. Daly, M. Burchett, F. Torpy, Plants in the Classroom Can Improve Student Performance, National Interior Plantscape Association, 2010, pp. 1–9.