World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:18, No:8, 2024

Detection of Cyberattacks on the Metaverse Based on First-Order Logic

Sulaiman Al Amro

***Abstract*—**There are currently considerable challenges concerning data security and privacy, particularly in relation to modern technologies. This includes the virtual world known as the Metaverse, which consists of a virtual space that integrates various technologies, and therefore susceptible to cyber threats such as malware, phishing, and identity theft. This has led recent studies to propose the development of Metaverse forensic frameworks and the integration of advanced technologies, including machine learning for intrusion detection and security. In this context, the application of first-order logic offers a formal and systematic approach to defining the conditions of cyberattacks, thereby contributing to the development of effective detection mechanisms. In addition, formalizing the rules and patterns of cyber threats has the potential to enhance the overall security posture of the Metaverse and thus the integrity and safety of this virtual environment. The current paper focuses on the primary actions employed by avatars for potential attacks, including Interval Temporal Logic (ITL) and behavior-based detection to detect an avatar's abnormal activities within the Metaverse. The research established that the proposed framework attained an accuracy of 92.307%, resulting in the experimental results demonstrating the efficacy of ITL, including its superior performance in addressing the threats posed by avatars within the Metaverse domain.

***Keywords*—**Cyberattacks, detection, first-order logic, Metaverse, privacy, security.

I. Introduction

THE advent of the Metaverse will initially enable access to online virtual spaces by means of augmented reality and virtual reality technologies. A primary feature of Metaverse technology concerns its prevalence and complexity, facilitated by an unprecedented fusion of the virtual and physical worlds [1]. In addition, rapid technological progress is expected for interfaces allowing interaction with the Metaverse, i.e. the technology and devices employed to input commands to, and receive feedback from, the Metaverse. Among these are brain-computer interfaces, such as neural interfaces designed to synthesize and process electrical signals taking place in the human brain because of certain cognitive activities, and convert them into meaningful input to a computer or external device (see Fig. 1).

The Metaverse will eventually be involved in conducting multisensory experiments and feedback, both through implants into the brain and alternative technology, including tactile devices, i.e. mechanical devices that mediate the process of communication between user and computer. For example, sensory systems will provide Metaverse users with Force return effects capable of simulating physical interactions in the real world, depending on the outcome of their actions in the three-dimensional virtual world [3]. This indicates that the Metaverse will be the next evolutionary step in the ability of human beings to offer and consume multimedia and multisensory content. However, this will inevitably lead to data security and privacy becoming one of the most pressing concerns related to Metaverse worlds, due to the sharing of high levels of personal information through virtual reality systems. This will therefore create a Data Hotspot capable of being targeted by cybercriminals [1]. In addition, data collection itself raises various issues, as companies will not just collect ordinary information (i.e., email addresses, usernames, and geographical location), but also biometric data and behavioral profiles, as well as additional personal information. The security issues in the Metaverse are shown in Fig. 2.

According to [4]-[6], companies will need to undertake actions above and beyond simple policy changes in order to protect users' data and privacy. It will be vital to create a reliable ecosystem able to build algorithms, frameworks, and regulations to address the issues related to privacy and security. This will be vital due to the potential for violations of privacy and security endangering the safety of interactions and users. An example is an allegation of sexual harassment against a participant in the beta version of the online virtual reality video game "Horizon worlds" (Horizon Worlds) owned by Meta (which represents one of the company's visions of the Metaverse). This has led to Meta introducing a tool called "personal boundaries" for users of Horizon worlds and Horizon Venues apps, to guarantee an ability to set a space of up to four feet between users' virtual avatars, in order to reduce incidents of virtual harassment and other abusive behavior. In addition, it is vital to protect a user's digital identity, in response to the danger of fraud and identity theft in the Metaverse. Moreover, the developers of Metaverse worlds, and in particular large companies, tend to collect personal information to serve their commercial and research purposes. Furthermore, many companies are also planning to broadcast advertisements in the Metaverse worlds, in order to take advantage of the interactions between users. This has led to the use of first-order logic to detect cyberattacks in the Metaverse becoming an emerging and critical area of research. The current study employed Interval Temporal Logic (ITL) as the logical framework, due to its appropriateness for delineating system traces. Specifically, ITL is effective in characterizing both desirable and undesirable behaviors. The Tempura tool was subsequently used to ascertain the presence of positive or negative behaviors by

Sulaiman Al Amro is with Department of Computer Science, College of Computer, Qassim University, Saudi Arabia (e-mail: samro@qu.edu.sa).

leveraging ITL descriptions with system traces.

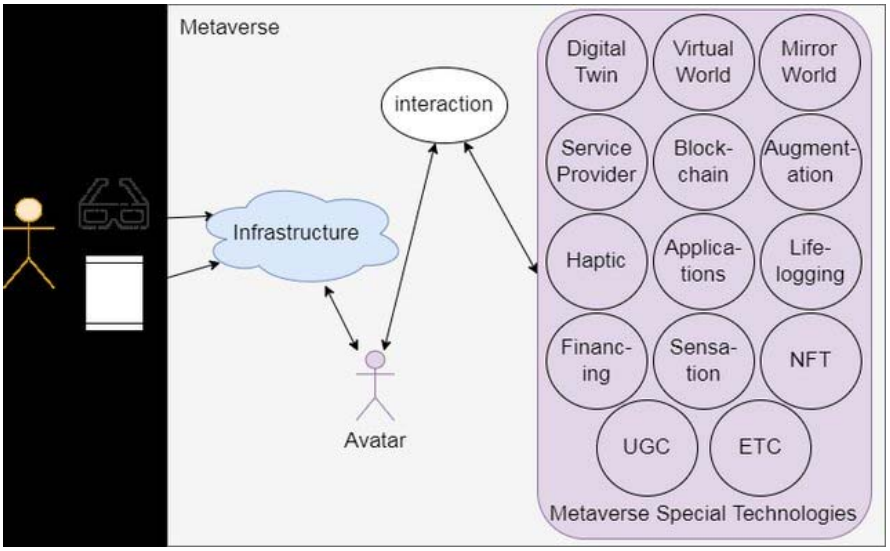

Fig. 1 Metaverse framework [2]

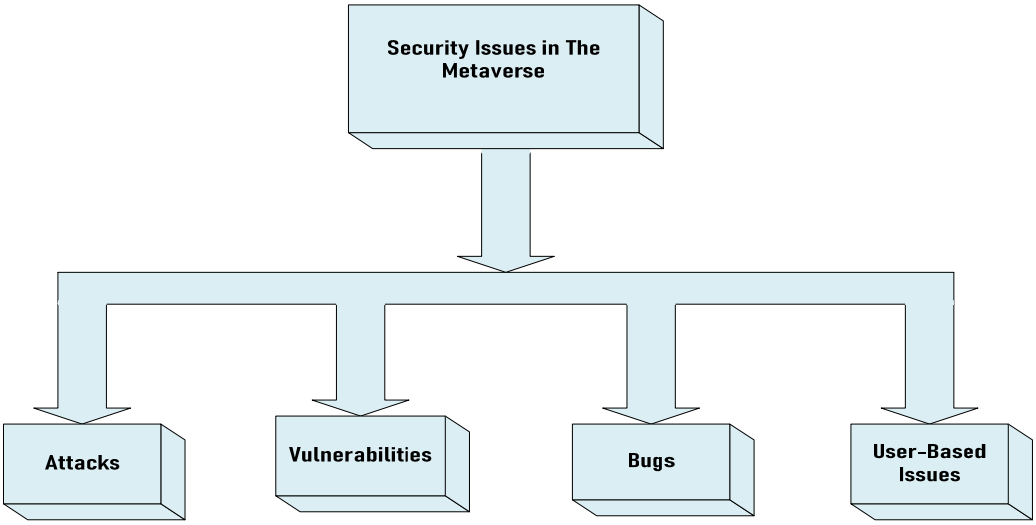

Fig. 2 Security Issues in the Metaverse [3]

II. Related Work

The research in [7] suggested utilizing recent developments in user-plane inference to enable the swift detection of (and response to) cyberattacks within the Metaverse, while at the same time maintaining a seamless user experience. The researchers emphasized the need to establish a rapid user-plane security framework, employing programmable switches and in-switch inference, to address the shortcomings of conventional controls and orchestrate plane-centric methods for detecting cyber threats within the Metaverse. The solution they suggested can be deployed in a practical programmable network testbed, highlighting its feasibility through the analysis of real-world network traffic data related to publicly accessible IoT-based threats. In addition, it employs performance metrics such as weighted F1 score, True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR), and False Negative Rate (FNR) to assess the ability to detect cyberattacks. Although the proposed solution appears to offer a promising method for detecting cyberattacks in the Metaverse, it has also been found to contain a number of limitations in relation to its features, scalability, model complexity, evaluation, and applicability.

In [8], a model constituted an inventive Intrusion Detection System (IDS) founded on deep learning principles. This specifically incorporated Kernel Principal Component Analysis (KPCA) for the extraction of attack features, and Convolutional Neural Networks (CNN) for the recognition and classification of such attacks. The primary objective of this model is the detection of diverse attacks directed at Metaverse-IoT (Internet of Things) communications. Its efficacy has been systematically assessed through the utilization of benchmark datasets encompassing potential IoT attacks directed towards

communication within the Metaverse. The outcomes have affirmed the efficacy of the proposed IDS model for discerning multiple classes of attacks pertinent to Metaverse-IoT communications. Additionally, its superior performance has been substantiated by the comparative analysis of the attack detection accuracy between the proposed IDS model and alternative techniques documented in the literature. However, an inherent limitation of the suggested model lies in the need for a cautious interpretation of its results, in response to the proposed IDS model undergoing testing and evaluation by means of benchmark IoT datasets. It has been suggested that these may not comprehensively encapsulate the complexities of real-world scenarios within Metaverse-IoT networks.

The research in [9] devised a method for human-avatar interaction capable of facilitating smooth engagement between multiple users within the Metaverse. This entails implementing an adaptive transmission strategy to share avatar motion information with other clients, as well as optimizing the data granularity of multiple avatars during network transmission. The researchers utilized a user study to assess the efficacy of the proposed solution, seeking to demonstrate technical feasibility, user acceptability, and enhancements in capturing human motion with heightened precision and reduced latency in comparison to preceding methodologies. In addition, [9] proposed a Human-Avatar Interaction Framework. This includes the creation of a three-dimensional representation of the user's avatar, equipped with a comprehensive skeletal structure designed to enable motion within the Metaverse. This framework utilizes Motion Capture (MoCap) software, exemplified by Axis Studio, to monitor the user's physical movements through wireless inertial sensors, i.e. those found in Perception Neuron. However, although the study did not explicitly highlight any limitations of the proposed framework, it should be noted that it was a preliminary evaluation, with further research being required to assess its scalability and performance in larger-scale environments and with more concurrent users. Additionally, the study only evaluated the framework's technical feasibility and user acceptability, and further research is needed to assess its impact on user experience and engagement in the Metaverse.

The model in [10] was proposed for real-time interactive avatar generation system in the document and revolves around the amalgamation of MediaPipe with the Unreal Engine. The operational procedure involves capturing real-time video input sourced from a webcam or pre-recorded videos. This subsequently undergoes processing through the MediaPipe plugin, which performs landmark estimation, encompassing facets such as facial expressions and body poses. The utilization of the MediaPipe Holistic model (an amalgamation of pose and face detection) contributes to the generation of precise and intricate avatar movements and facial expressions. Integration between the avatar in the Metaverse and the MediaPipe plugin is achieved through an animation blueprint, thus facilitating the avatar's synchronization with real-time input derived from the plugin. This model is instrumental in producing avatars capable of faithfully replicating the movements and expressions of human-beings in real-time. On the other hand, despite the proposed model providing the generation of real-time motion and expression for Metaverse avatars, it exhibits limitations pertaining to the intricacy of facial expressions, alongside a reliance on video input, performance considerations, and integration with additional features, as well as compatibility with various platforms.

III. Main Framework

This current research explored the behavior of avatars within the Metaverse. This involved a close examination of their actions, to gain insights into how they operate within the system, and thus, uncovering their patterns of behavior. Avatars' activities were systematically monitored as a series of sequential steps to trace their movements. When an avatar intends to infiltrate a system, it follows a set of discrete steps, collectively forming its unique behavior, and which serves as the foundation for defining specifications that align with our initial system requirements (see Fig. 3). These specifications serve as the basis for generating formulas that can be deduced from the analysis of avatar behavior.

IV. VR Testing Environment

We set up our experimental environment using the XR Interaction Toolkit to control the mechanics and implement diverse interaction techniques. We utilized a Room-Scale XR Rig [5], which grants users the capability to move freely in six degrees of freedom. The virtual setting was deliberately designed to mimic an office space, with objects placed in random arrangements allowing users to interact as they pleased. In addition, we used two distinct desktops, each equipped with separate Microsoft Windows operating systems to scrutinize the actions of the avatars. Both computers possessed unrestricted Internet access and featured contemporary, widely used desktop and Internet applications. The inclusion of these applications in the normal action testing served the dual purpose of assessing usability as a real-time monitor and detector. Furthermore, we ensured that, throughout testing, antivirus software was active on both desktops, so as to evaluate the system's efficacy in the presence of concurrent detectors.

V. Analysis of Avatar Behavior

The virtual analysis of avatar behavior confirmed that, in order to harm a system, avatars tend to focus on four main actions: (1) fraud; (2) crime; (3) defamation; and (4) Identity theft. We therefore scrutinized the steps of an avatar's behavior within our system by observing its methods in real-time. This approach involved examining these actions to determine the presence of any potentially malicious activities, in order to assess indications of malicious behavior.

We began by examining the Rootkit dropper, the module tasked with loading (dropping) the rootkit. In the context of an avatar, all loading occurs in the memory. Consequently, the dropper requires additional effort to load DLL modules and the rootkit. In this research, Avatar Behavior Analysis determined that an avatar would attempt the following three main actions to attack the Metaverse: Step (1) Concealment: Within the

virtual Metaverse, a malevolent actor, adopting the guise of a malicious avatar, deliberately conceals their identity, in order to mimic the appearance of the target avatar, thereby deceiving the interacting individuals. Step (2) Personification: In the corporeal domain, the assailant (functioning as a malicious manipulator) procures authentication for a device through a legitimate player, subsequently assuming his/her identity, so as to manipulate the corresponding avatar. Step (3) Replication: In both the Metaverse and the physical realm, the aggressor gathers obsolete identity parameters that are linked to an honest avatar before, falsely asserting to be the target avatar, and presenting them to the interactor.

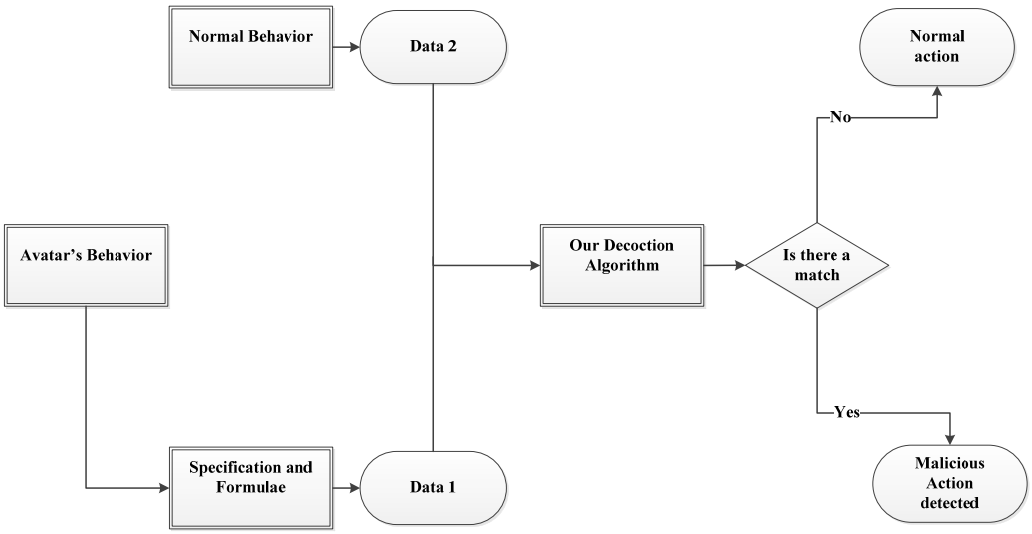

Fig. 3 The main framework

VI. Avatar Detection System

In this research, our selection of this logical framework was motivated by the presence of Tempura, an executable subset of ITL facilitating the fulfillment of our system specifications [11]. Moreover, ITL proved highly suitable for delineating system traces, effectively capturing both undesirable and favorable behaviors [6]. We officially designated step 1, step 2, and step 3, denoting the comprehensive actions for Concealment, Personification, and Replication. We suppose that the variable X signifies the entirety of avatar actions obtained during runtime within the Metaverse. X is expected to be received in textual form, encapsulating the representation of all actions. Note that for a given set A, the predicate 'in A(X)' holds if X ϵ A.

For step one, the ITL formulae will be as follows:

$$Ustep1(X) = (inmeta\ (X)) \wedge instep1(X) \quad (1)$$

For step two, the ITL formulae will be as follows:

$$Ustep2(X) = (inmeta\ (X)) \wedge instep2(X) \quad (2)$$

For step three, the ITL formulae will be as follows:

$$Ustep3(X) = (inmeta\ (X)) \wedge instep3(X) \quad (3)$$

Therefore, the subsequent formula is utilized to catch any suspicious actions by the avatar as shown in (4):

$$\Upsilon\ (inmeta\ (X) \wedge (Ustep1(X) \vee Ustep2(X) \vee step3(X)) \quad (4)$$

The previous formula indicates that the examination will be undertaken if the avatar is in the Metaverse, and has completed one (or more) steps. In addition, the order of these steps is significant.

The previous formula suggests that the occurrence of an attack by the avatar is indicated when one or more steps are detected. Furthermore, the specific order of these steps is highly significant and constitutes the primary contribution made by this research. The formulas are designed to check whether certain conditions are met during the runtime of the Metaverse.

A. Steps' Formulas

Ustep1(X), Ustep2(X), Ustep3(X) are defined based on the conditions of being in the Metaverse and satisfying specific step conditions.

B. Suspicious Actions Formula

The formula *ϒ (inmeta(X) ʌ (Ustep1(X) ᵛ Ustep2(X) ᵛ Ustep3(X)))* checks for suspicious actions. It verifies that the avatar is in the Metaverse and has performed one or more steps in a specific order.

C. Attack Scenarios

The subsequent formulas (5 to 10) represent differing sequences of actions the avatar may undertake to launch an attack. These scenarios cover various orders of executing actions from the three steps.

$$\Diamond\ Ustep1(X); \Diamond\ Ustep2(X); \Diamond\ Ustep3(X) \quad (5)$$

$$\Diamond\ Ustep1(X); \Diamond\ Ustep3(X); \Diamond\ Ustep2(X) \quad (6)$$

$$\Diamond\ Ustep2(X); \Diamond\ Ustep1(X); \Diamond\ Ustep3(X) \quad (7)$$

$$\Diamond\ Ustep2(X); \Diamond\ Ustep3(X); \Diamond\ Ustep1(X) \quad (8)$$

$$\Diamond\ Ustep3(X); \Diamond\ Ustep1(X); \Diamond\ Ustep2(X) \quad (9)$$

$$\Diamond\ Ustep3(X); \Diamond\ Ustep2(X); \Diamond\ Ustep1(X) \quad (10)$$

The previous six formulas demonstrate that, in order to attack, an avatar will undertake different actions from the three steps in six orders. Firstly, the normal order from step one to step three. Secondly, the avatar will execute activity from step one, three and then two, respectively. Consequently, the avatar will firstly do actions from step two then step one, and subsequently actions from step three. The fourth scenario will be step two, three, and one. Hereafter, the avatar will first execute activity from step three, then step one, and subsequently step two. The final scenario is that the avatar initially undertakes actions from step three, then step two, and lastly step one.

A Java program served as an intermediary between Tempura and the Avatar action, facilitating the conversion of the output from malicious actions into a format readable by Tempura. The Java program functioned as a conduit, ensuring seamless communication between Tempura and the action behavior.

Tempura obtained the three previously described steps from the system as distinct lists, with each list being formatted as a string list for Tempura's capture. Whenever an avatar executes, it promptly furnishes all relevant information to the Java program, which functions as an intermediary conduit. Simultaneously, Tempura reads the information from the Java program.

To validate the theory, we tested a set of typical avatar actions to demonstrate that these normal behaviors do not execute steps identical to those that are malicious. We conducted two experiments to examine malicious actions: firstly, malicious analysis, utilizing established tools to define the behavioral steps of an avatar, as previously discussed, and secondly, the detection of malicious actions. The outcomes of the latter experiment are detailed in the following subsection.

The research prototype is capable of concurrently analyzing multiple actions. However, the testing of avatar actions was performed individually. Consequently, malicious actions were assessed sequentially, with the virtual machine being restored after each test. We adopted this approach to guarantee a pristine operating environment for each attack, minimizing the possibility of interference between various malicious actions.

The outcomes of the analysis of avatar actions can be considered insignificant and unnecessary for inclusion in the comprehensive results, but their exclusion would leave the theoretical objectives of this research unfulfilled. Table I illustrates the results of the analysis of 12 malicious actions associated with the three steps, i.e. three actions in each step representing malicious avatar actions. Fig. 4 shows that 12 out of 13 attempts were detected, i.e., approximately 92%.

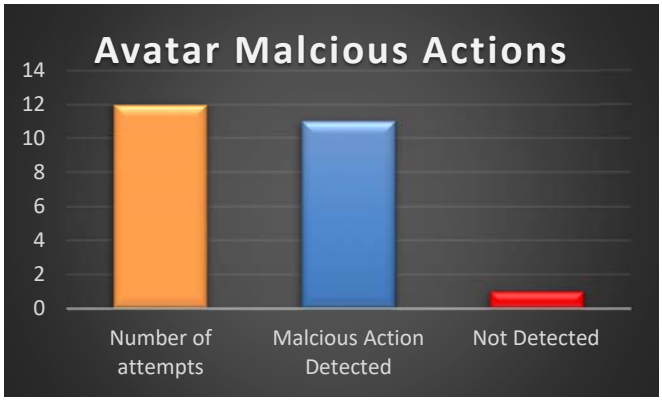

Fig. 4 Results of the avatar analysis

TABLE I
MALICIOUS AVATAR ACTIVITIES

Number	Attempt action	Category	Detected or not
1	Scrutinize the publicly accessible components of others' online identities to acquire relevant information about them.	Step 1	Detected
2	Search for other's sensitive information, like payment card details.	Step 1	Detected
3	Hiding one's real identity or using fake profiles.	Step 1	Detected
4	Search for open ports on the machine.	Step 2	Detected
5	Seek to acquire valuable topological insights into another entity's internal network using DNS zone transfer.	Step 2	Detected
6	Attempt to induce alterations in the environment, such as introducing new elements or modifying existing ones.	Step 3	Detected
7	Redirecting to malicious content or an exploit kit by exploiting cross-site scripting.	Step 2	Detected
8	Seek to manipulate the values of specific variables with intent.	Step 3	Not Detected
9	Attempt to transmit deceptive communications with the aim of compelling a victim to disclose classified credentials or other sensitive information.	Step 1	Detected
10	Try to exploit misconfiguration to infiltrate another network.	Step 2	Detected
11	Seek to establish supplementary user accounts to retain access to a system.	Step 3	Detected
12	Concentrate on identifying and leveraging weaknesses in authentication mechanisms.	Step 3	Detected
13	Attempt to impair or obstruct the availability of services for others.	Step 2	Detected

These results signify that a sizable portion of malicious actions, across various classes, were successfully integrated into the system. However, certain actions deviated from the theoretical framework of this research by failing to adhere to this integration. This indicates that these malicious actions did not conform to the three specified steps upon which this research is grounded. Two potential explanations are posited. Firstly, these actions may have been unsuccessful in locating a suitable environment or victim file for attachment. This is due to an action requiring external support to propagate, i.e. a system service, a file, or another form of assistance. The absence of such aid results in a failure of the action to infect the system. Alternatively, the codes may completely abstain from

attaching themselves to other files, so rendering them incapable of being classified as malicious actions.

VII. Conclusion and Future Research

This research concludes that, to secure a dependable environment for users, a number of considerations need to be taken into account during the creation of Metaverse worlds. Simultaneously, it is vital that the success and sustainability of this nascent technology is also safeguarded. This paper presents a comprehensive framework designed to specify, implement, and validate a behavior-based avatar detection system within the Metaverse. The study explored a method of identifying a distinctive characteristic inherent in certain avatar actions. This characteristic was initially formalized using ITL and subsequently translated into Tempura programming for its practical implementation. Consequently, we consider that the proposed approach is capable of effectively identifying any attempts by an avatar to compromise the integrity of the system.

Our current research entails the execution of a comprehensive experiment to both validate and assess the threat posed by avatars using AnaTempura, an integrated workbench tool designed for ITL that aligns with our system specifications. We consider that AnaTempura will play a crucial role in validating and verifying the ITL specifications, along with facilitating runtime testing of these specifications. Furthermore, we conclude that the inherent complexity of computer systems will continue to prompt researchers to explore alternative solutions. This therefore indicates the need for future research to further enhance any threat detection approach in this context.

References

[1] K. Yang, Z. Zhang, Y. Tian. and J. Ma. "A secure authentication framework to guarantee the traceability of avatars in Metaverse." IEEE Transactions on Information Forensics and Security. 2023.

[2] N. A. Dahan, M. Al-Razgan, A. Al-Laith, M. A. Alsoufi, M. S. Al-Asaly, and T. Alfakih, "Metaverse framework: A case study on E-learning environment (ELEM)." Electronics, 11(10), p.1616. 2022

[3] R. Di Pietro and S. Cresci. "Metaverse: security and privacy issues." In 2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA) (pp. 281-288). Dec 2021

[4] U. Kose. "Security Issues in Artificial Intelligence Use for Metaverse and Digital Twin Setups". In: E. Karaarslan, Ö. Aydin, Ü Cali, M. Challenger, (eds) Digital Twin Driven Intelligent Systems and Emerging Metaverse. Springer, Singapore. https://doi.org/10.1007/978-981-99-0252-1_17 2023.

[5] Unity Technologies, https://learn.unity.com/tutorial/configuring-an-xr-rig-with-the-xr-interaction-toolkit. Accessed 03/08/2023.

[6] S. Al Amro, and A. Cau. "Behaviour-based virus detection system using Interval Temporal Logic." In 2011 6th International Conference on Risks and Security of Internet and Systems (CRiSIS) (pp. 1-6). IEEE. Sept 2011.

[7] B. Bütün, A. T. J. Akem, M. Gucciardo, and M. Fiore. "Fast Detection of Cyberattacks on the Metaverse through User-plane Inference." In International Conference on Metaverse Computing, Networking and Applications. June 2023.

[8] T. Gaber, J. B. Awotunde, M. Torky, S. A. Ajagbe, M. Hammoudeh and W. Li. Metaverse -IDS: "Deep learning-based intrusion detection system for Metaverse -IoT networks." Internet of Things, 24, p.100977. 2023.

[9] K. Y. Lam, L. Yang, A. Alhilal, L. H. Lee, G. Tyson, and P. Hui, P. Human-avatar interaction in Metaverse: Framework for full-body interaction. In Proceedings of the 4th ACM International Conference on Multimedia in Asia (pp. 1-7). Dec 2022.

[10] E. A. Tuli, A. Zainudin, M. J. A. Shanto, J. M. Lee, and D. S. Kim. "MediaPipe-based Real-time Interactive Avatar Generation for Metaverse." InProc. Korea Commun. Soc. Conf., pp.1370-1371. 2023.

[11] J. Y. Halpern and B. C. Moszkowski. "Executing temporal logic programs". Cambridge University Press, Cambridge etc. 1986, xiii+ 125 pp. Journal of Symbolic Logic, 53(1). 1988.

Sulaiman Al Amro is currently working as an Associate Professor in computer science department at Qassim University. He received a BS in computer science from Qassim University in 2006, and master in IT De Montfort University (DMU), Leicester (UK) in 2009, and PhD in computer science from De Montfort University (DMU), Leicester (UK) in 2013. He has been involved in several program committees and is being a Reviewer in different international conferences and journals. Dr. Sulaiman has a number of research interests including Cybersecurity, Software Engineering and Artificial Intelligence.