Security Vulnerability Analysis in Deploying Avatar in the Metaverse

Arpita Dinesh Sarang
SysCore Lab.
(Convergence Engineering for Intelligent Drone)
Sejong University
Seoul, Republic of Korea
arpitasarang98@gmail.com

Ki-Woong Park*
Dept. of Computer and Information Security
Sejong University
Seoul, Republic of Korea
woongbak@sejong.ac.kr

***Abstract*— Cyberattacks have affected every technology in recent years, and the metaverse is not an exception. Users can engage with other users by navigating the metaverse with their avatar's associated user data. Attackers using the user data associated with avatars masquerade as users or penetrate through the metaverse's vulnerable infrastructure. There are multiple components of the metaverse infrastructure that are vulnerable to attack. The infrastructure becomes insecure when these components are developed without security in consideration. Cyberattacks that take advantage of these vulnerabilities compromise user privacy and cause the loss of sensitive data. This highlights the necessity of comprehending the infrastructure and the data flow through its essential components so that appropriate countermeasures can be built. Our study proposed a model that clarifies the data flow among the infrastructure's components for creating an avatar in the metaverse. We also identify the major threat causing cyberattacks per component. We state in our study's conclusion that such threats will have an impact on the metaverse's avatar and user data.**

Keywords—avatar, metaverse, cybersecurity, cloud computing, IoT (Internet of Things), Artificial Intelligence(AI)

I. INTRODUCTION

The Metaverse is a virtual reality environment where individuals engage with each other and a constantly evolving computer-generated environment through Avatars. User data is linked to each avatar, which is a distinctive digital representation of the user. The Metaverse is evolving in the direction of appearing almost real. Every day, a new theory to enhance the metaverse is developed. An individual can use devices at home, on the go, at work, or in public spaces to access the metaverse. These devices store information about the user, such as their location, age, shopping preferences, friends, favorite movies, mother's name, credit card number, bank information, medical information, social security number, and so forth [1]. Since the metaverse involves an individual's sensitive information and identity, each theory for evolving metaverse should be developed with a robust security plan. The privacy of personal information, behavior, and communications should be retained belonging to the user's avatar in the metaverse. Therefore, many threats exist for the metaverse when it comes to network communication, access control, data management, authentication, and sharing of data with third-party organizations [2]. Additionally, both the metaverse and avatars are software, the metaverse is also a multitude of diverse applications that share user data. Thus, building API automation for data sharing is inevitable [3]. User data is handled by cloud computing, edge computing, and other computing paradigms that have policies and professionals involved [4]. This endangers the metaverse, humans and avatars, avatar behaviour, the metaverse's technological capabilities, and the metaverse's outputs [5]. This is where the data flow and security risk in the metaverse need to be assessed. Our suggested model identifies the components that are susceptible to cyberattacks as well as the data flow. This will support the development of more reliable programs for the cloud, devices, network communications, metaverse software, and AI models.

* Corresponding author

II. RELATED WORK

A variety of studies of how the Metaverse is vulnerable to cyberattacks have lately been published. Artificial intelligence (AI) technology makes up the metaverse. They rely on edge computing (EC), 5G infrastructure, and federal learning (FL) frameworks. Using famous attacks like phishing, man-in-the-middle, denial-of-service attacks, SQL injections, zero-day attacks, and DNS tunnelling, unauthorized access to a data on computer, network, or devices connected to the network can be destroyed or disrupted [6]. Developing cyber security strategies is necessary to protect against physical threats, privacy issues, social issues, and identity theft when gaining access to the metaverse. XR authentication, AI-driven cybersecurity, access control policies, cybersickness mitigation, and XR forensics are described as alternatives to traditional methods for these issues [7].

Lack of security planning led to a new attack, the Man-in-the-Room (MitR), from a vulnerable social networking application that was made feasible by the unique features of the virtual reality landscape. Worming and botnet capabilities were modified for VR, potentially having a major impact on millions of users [8]. Future developments of diversified, virtual, and more sophisticated networks will result from the fusion of the metaverse with the Internet of Things (IoT). Therefore, there is a need of innovative Deep learning-based Intrusion Detection Systems (IDS) models in order to detect the majority of assaults targeting metaverse-IoT connections [9]. However, previous studies did not clarify which, why, or where the cyberattacks in the Metaverse architecture were targeted. The Metaverse architecture and the cyberattacks that target its various components are justified by our proposed model.

III. PROPOSED MODEL

The metaverse is a virtual world where digital representations of real-world objects and people can coexist.

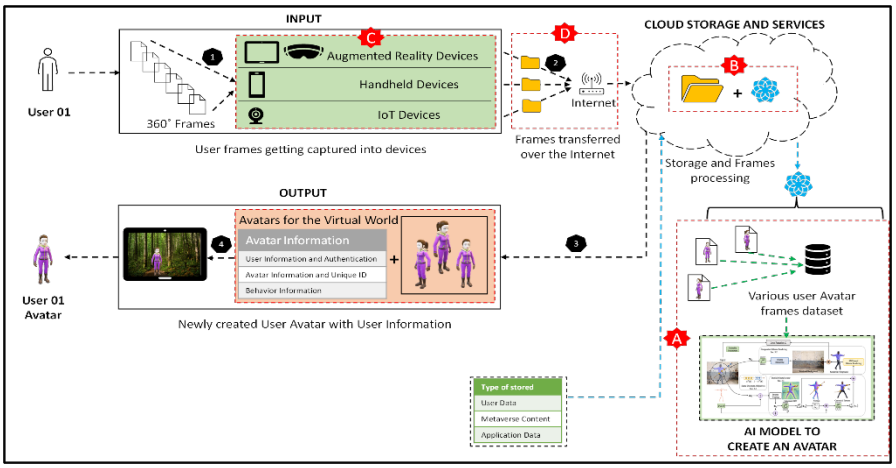

Fig. 1. SARANG (Simplified Avatar Relationship Association with Non-linear Gradient) Model

Our model demonstrates how a person is captured with a camera, processed into a digital avatar of that person, and then saved in cloud storage along with all of their unique characteristics. AI models are used to process the recorded human frames and create a distinct avatar with a unique identity. Through the internet, the unique identity and real-time behavior of the avatar are transmitted and updated in the cloud storage. According to the SARANG (Simplified Avatar Relationship Association with Non-linear Gradient) Model, the data flow is as follows: the input is taken as frames, processed using an AI model to create a unique avatar, and then simultaneously stored in cloud storage. The data flow via the model is defined by points 1 to 4. However, the components of the dataflow architecture that are vulnerable are listed in points A, B, C, and D, in Figure 1.

Firstly, to access and navigate a metaverse environment, an individual must first construct their avatar which is a distinctive digital representation. The user will use a handheld device, an AR (augmented reality) device, or an Internet of Things (IoT) device with cameras and sensors to construct avatars. The metaverse software on these devices will record numerous image frames 360 degrees for input. Secondly, the input is then sent over the internet for storage over the 5G or 6G network. Utilizing cloud computing, edge computing, and other computing paradigms, storage stores user input as raw or initial data on cloud storage provided by third-party vendors. The AI model required to process the input is likewise stored in the cloud. As a result, the AI model and the saved input data combine to create an avatar, which is then given its own user information, authentication, unique ID, and behavior data. Finally, the user receives the newly constructed avatar together with its information transmitted via the internet. At that point, Avatar is prepared to explore the metaverse. This model clarified that AI models, cloud computing, IoT devices, and networks are the four primary components involved in the creation, processing, storing, and transmission of data associated with Avatar. These elements are all susceptible to cyberattacks.

IV. THREAT IDENTIFICATION ON THE MODEL

In the proposed SARANG model, we try to identify the threats while communicating AI Models, Cloud storage, IoT Devices, and Networks and the impact of cyber-attacks on them through our study.

A. AI Model(component A)

In the proposed model we use the AI model as an example model which can be replaced by any model in use for Avatar creation. It is evident that reconstructing 3D humans from videos of them in the wild is difficult. Its solution necessitates precisely separating people from random backgrounds. The objective of the included AI model is to be able to precisely reconstruct intricate clothing deformations and unique facial features from monocular videos. Innovative techniques for 2D segmentation, unique view synthesis, and reconstruction challenges are used to assess the model. The training is designed as global optimisation to simultaneously optimise the per-frame pose parameters, and the dynamic foreground and static background fields[10]. AI models pose serious risks to the security and privacy of Smart Cities since they are more prone to cyber-attacks. Such as, **Case 1:** According to Facebook, one can "connect, work, play, learn, and shop" in the metaverse. This facilitates activity completion, business dealings, and virtual presence at necessary locations, all of which contribute to virtual human interaction. Each event generates data, and the data about the human avatar is linked to that data. Digital twins, deepfakes, or interrupting an avatar's metaverse participation might cause miscommunication and generate inaccurate data that could be adverse to both the organisation and the metaverse users. **Case 2:** Despite certain obstacles, medical diagnosis, patient monitoring, medical education, surgeries, and medical therapies could all benefit from the use of the metaverse in the field of medicine. At Seoul National University Bundang Hospital in South Korea, advanced training in lung cancer surgery was delivered through the use of XR, VR, and AR technology [11]. The use of these to enable practitioners to be virtually accessible facilitates the sharing of important data and expertise, but it also poses information security threats in the event that the avatar's data is misconstrued and endangers the indigenous knowledge that will be further practised in reality. Furthermore, laws and guidelines about the security of user data are being implemented globally. Federated learning (FL), in contrast to centralized machine learning, offers a natural way to preserve users' privacy by dispersing learning over decentralized still technological issues with it, and there are known risks [12]. When implementing AI models, there are a number of challenges to overcome as well as opportunities for defense, including model watermarking, information hiding problems and defensive strategies, adversarial learning and model resilience, and models with fairness considerations [13]. A recent study suggests cyberattacks based on artificial intelligence and analyzes them to determine appropriate cyber defenses [14]. According to our analysis, the major three cyberattacks that have the most effects on AI models:

- Model threat

  In the case of the attackers have the ability to either access the AI model or source code that generates the Avatar. Once attackers have this model, they can use it to study how it reacts to different inputs, reverse engineer it, which is also known as a model inversion attack, and then develop malicious prompts by figuring out its vulnerabilities. An attacker may be able to control, malfunction, or record Avatar activities in the metaverse.

- Data Poisoning

  Large-scale data analysis is the first step in the process of using machine learning and AI models to identify patterns and generate predictions for Avatar-related data and Avatar Behavior. Attacks can get more creative due to this core process. Malicious training data can be injected by attackers to teach the avatar-generating AI models false information, which leads to inaccurate, dishonest, or malicious activities. The integrity of the data in the metaverse is improved by assessing data sources and flows and end-to-end certifying training processes with blockchain, reliable hardware, and formal verification.

- Backdoor AI

  Another technique attackers employ to alter or update an AI system is an AI backdoor Model. Assuming that attackers can access the server that stores the models that generate Avatars. They upload a trojan model, which is trained on a different kind of data but looks exactly the same. This leads to a serious security disaster for Avatars. For example: it won't restrict the use of offensive words. In order to overcome the corrupted inputs, we need to either eliminate them by retraining the model or retain them and implement fine-tuning. The model should be secured by rejecting adversarial input and then issuing an alert.

B. Cloud Computing (component B)

Organizations and individual users move their applications, data, and services to the cloud storage server because of its scalability and availability for computing activities. Despite its benefits, the shift from local to remote computing has presented a number of security risks and difficulties for both service providers and customers [15]. It was essential to state security issues in other relevant fields, such as trust-based security models, cloud-enabled Big Data applications, the Internet of Things (IoT), Software Defined Networks (SDN), and Network Function Virtualization (NFV) [16]. It is difficult to determine where the data is kept because cloud providers do not disclose the location of the data only accessed via the internet [17]. This makes cloud storage prone to the following cyber-attacks:

- Denial of Services

  An attempt to prevent authorized users from accessing their avatar or the metaverse environment data being stored on the cloud is known as a denial-of-service (DoS) attack. DoS attacks usually involve sending a lot of traffic to a cloud service at once, overloading it. DoS attacks can have an adverse impact on an organization's reputation by impairing its capacity to provide essential services and resulting in monetary losses. Due to the size and complexity of cloud settings, cloud-based DoS assaults can be very difficult to protect against in terms of attack identification and mitigation. Keeping track of network traffic and safeguarding data while it is in transit and backup storage, the provider must be willing to submit information about external audits, security certifications, and hash and encryption techniques, as well as key lengths.

- Insecure APIs

  Vulnerabilities in APIs communicating with the cloud, AI Model, and Avatar allow attackers to access systems or data without authorization or to interfere with the API's operation. APIs are flawed in two ways: Shadow APIs: APIs that are not properly permitted or documented, and the unknown entity that owns the API being unaware. These APIs may be made by developers or other professionals which may provide unauthorized individual access to private information. API parameters: The inputs and outputs of an API should be properly validated and filtered, as they are susceptible to injection attacks.

- Security Misconfiguration

  When cloud computing communicating resources and data flow infrastructure for Avatar and the metaverse are improperly configured to defend against cyberattacks, this is known as security misconfiguration. This can involve not configuring and securing systems and software appropriately, setting access controls inappropriately, and failing to update and patch systems and apps regularly.

C. IoT Devices (component C)

Preserving privacy and confidentiality, ensuring the security of users, infrastructures, data, and IoT devices, and ensuring the availability of services provided by an IoT ecosystem are the primary goals of IoT security [18]. The IoT's current and future applications hold immense potential for improving user comfort, productivity, and automation levels. Therefore, architecture upgrades are required to achieve end-to-end secure IoT environments [19]. We discuss the major security risks associated with IoT device cyber-attack:

- Device Spoofing

  A kind of attack where a malicious user Avatar impersonates a legitimate one by altering the IP address, MAC address, or other identifying information of an authentic device enters the metaverse. Network switches, setting port security and updating firmware on a regular basis can prevent this attack.

- Man in the Middle

  The idea behind a Man in the Middle (MitM) attack is for a hacker to eavesdrop on two Avatars' conversations. The attacker pretends to be the original Avatar as they are receiving trustworthy private data and interfere with services. Avoid this by accessing your avatar in the metaverse via a secure network.

- Zero-Day Attack

  In a zero-day attack, a hacker makes use of unpatched metaverse software vulnerabilities in Internet of Things devices that cybersecurity engineers were previously unaware of and not prevention available. The device's software must be updated.

D. Network Analysis (component D)

Network security is being seriously compromised by the growing expertise of attackers and their capacity to take advantage of software and firmware flaws. However, a lot of businesses frequently overlook the essential precautions needed to defend networks [20]. Attacks using networks to cause the clocks to desynchronize demonstrate attacks with little resources [21]. Such minor ignorance causes network-based attacks that may affect the Avatar as follows:

- Unauthorized Access

  An attacker who gains access to a network without authorization is considered to be using unauthorized access in our model's data flow. Weak passwords, inadequate protection against social engineering, prior compromised accounts, and insider threats are a few of the reasons why unauthorized access attacks occur.

- Code and SQL attacks

  Many websites use user input without properly validating and filtering it. Following that, attackers can submit malicious code in place of the anticipated data values while completing Avatar creation or initiating an API call. Attackers can compromise the cloud and devices connected by executing the code on it.

- Privilege Escalation

  After penetrating the network, attackers might utilize privilege escalation for Avatars in the metaverse to gain more access within the security perimeter. Attackers can obtain access to more systems by using horizontal privilege escalation, and escalate their privileges vertically to obtain higher access to the same systems. An excellent service for access control management ought to be included.

Threat identification for the SARANG model highlights significant attacks on Avatar creating dataflow through the infrastructure's components that have the potential to compromise the data's accessibility, confidentiality, and integrity. Data theft, abuse, and manipulation impact both the user and their avatar and degrade the reputation of the company that owns the metaverse.

V. Conclusion

From this study, we will have an understanding of how avatar generation data flow works using our proposed SARANG model, which consists of four components: AI models, cloud storage, IoT devices, and networks that represent a risk of cyber security. We discuss how much of an impact cyberattacks can have on these components and how vulnerable they are to cyberattacks. A cyber-attack on the SARANG model data flow will have a severe impact on the privacy of user data and lead to the malfunctioning of avatars in the metaverse. Hence, this prepares us for developing robust programs and countermeasures against cyber-attacks for the security of Avatars in the metaverse.

Acknowledgment

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) (Project No. RS-2022-00165794, 30%; Project No. 2022-0-00701, 10%; Project No. RS-2023-00228996, 10%), the ICT R&D Program of MSIT/IITP (Project No. 2021-0-01816, 10%), and a National Research Foundation of Korea (NRF) grant funded by the Korean government (Project No. RS-2023-00208460, 40%).

References

[1] Falchuk, B., Loeb, S. and Neff, R., 2018. The social metaverse: Battle for privacy. IEEE Technology and Society Magazine, 37(2), pp.52-61.

[2] Wang, Y., Su, Z., Zhang, N., Xing, R., Liu, D., Luan, T.H. and Shen, X., 2022. A survey on metaverse: Fundamentals, security, and privacy. IEEE Communications Surveys & Tutorials.

[3] Di Pietro, R. and Cresci, S., 2021, December. 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). IEEE.

[4] Ning, H., Wang, H., Lin, Y., Wang, W., Dhelim, S., Farha, F., Ding, J. and Daneshmand, M., 2023. A Survey on the Metaverse: The State-of-the-Art, Technologies, Applications, and Challenges. IEEE Internet of Things Journal.

[5] Narin, N.G., 2021. A content analysis of the metaverse articles. Journal of Metaverse, 1(1), pp.17-24.

[6] Pooyandeh, M., Han, K.J. and Sohn, I., 2022. Cybersecurity in the AI-Based metaverse: A survey. Applied Sciences, 12(24), p.12993.

[7] Chow, Y.W., Susilo, W., Li, Y., Li, N. and Nguyen, C., 2022. Visualization and Cybersecurity in the Metaverse: A Survey. Journal of Imaging, 9(1), p.11.

[8] Vondráček, M., Baggili, I., Casey, P. and Mekni, M., 2023. Rise of the metaverse's immersive virtual reality malware and the man-in-the-room attack & defenses. Computers & Security, 127, p.102923.

[9] Gaber, T., Awotunde, J.B., Torky, M., Ajagbe, S.A., Hammoudeh, M. and Li, W., 2023. Metaverse-IDS: Deep Learning-based Intrusion Detection System for Metaverse-IoT Networks. Internet of Things, p.100977.

[10] Guo, C., Jiang, T., Chen, X., Song, J. and Hilliges, O., 2023. Vid2avatar: 3d avatar reconstruction from videos in the wild via self-supervised scene decomposition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12858-12868).

[11] Chengoden, R., Victor, N., Huynh-The, T., Yenduri, G., Jhaveri, R.H., Alazab, M., Bhattacharya, S., Hegde, P., Maddikunta, P.K.R. and Gadekallu, T.R., 2023. Metaverse for healthcare: A survey on potential applications, challenges and future directions. IEEE Access.

[12] Rasha, A.H., Li, T., Huang, W., Gu, J. and Li, C., 2023. Federated learning in smart cities: Privacy and security survey. Information Sciences.

[13] Caviglione, L., Comito, C., Guarascio, M. and Manco, G., 2023. Emerging challenges and perspectives in Deep Learning model security: A brief survey. Systems and Soft Computing, p.200050.

[14] de Azambuja, A.J.G., Plesker, C., Schützer, K., Anderl, R., Schleich, B. and Almeida, V.R., 2023. Artificial Intelligence-Based Cyber Security in the Context of Industry 4.0—A Survey. Electronics, 12(8), p.1920.

[15] Singh, A. and Chatterjee, K., 2017. Cloud security issues and challenges: A survey. Journal of Network and Computer Applications, 79, pp.88-115.

[16] Kumar, R. and Goyal, R., 2019. On cloud security requirements, threats, vulnerabilities and countermeasures: A survey. Computer Science Review, 33, pp.1-48.

[17] Behl, A., 2011, December. Emerging security challenges in cloud computing: An insight to cloud security challenges and their mitigation. In 2011 World Congress on Information and Communication Technologies (pp. 217-222). IEEE.

[18] Hassan, W.H., 2019. Current research on Internet of Things (IoT) security: A survey. Computer networks, 148, pp.283-294.

[19] Hassija, V., Chamola, V., Saxena, V., Jain, D., Goyal, P. and Sikdar, B., 2019. A survey on IoT security: application areas, security threats, and solution architectures. IEEE Access, 7, pp.82721-82743.

[20] Arogundade, O.R., 2023. Network Security Concepts, Dangers, and Defense Best Practical. Computer Engineering and Intelligent Systems, 14(2).

[21] Berardi, D., Tippenhauer, N.O., Melis, A., Prandini, M. and Callegati, F., 2023. Time sensitive networking security: issues of precision time protocol and its implementation. Cybersecurity, 6(1), pp.1-13.