Virtual Sessions for Forensic Analysis of Video Conferencing Systems: A Novel Methodology

Jaykumar Soni, Tom Neubert, Benjamin Dietrich, Claus Vielhauer
Department of Informatics & Media
Brandenburg University of Applied Sciences
Brandenburg, Germany
{jaykumar.soni, tom.neubert, benjamin.dietrich, claus.vielhauer}@th-brandenburg.de

***Abstract*—During the last years online meetings and consequently Video Conferencing Systems (VCS) have become more and more popular to reduce travel time and costs. Because of the dramatically increased usage of VCS, it has become more important to analyze and evaluate their security and privacy due to the huge amount of privacy related multimedia data processed by VCS. Thus, in this paper, we present a novel privacy preserving methodology to generate virtual video conferencing sessions with reproducible data to enable a comparable and reliable analysis of these systems in future work.**

***Index Terms*—Privacy, Video Conferencing Systems, Media Security, Forensic Analysis**

I. Introduction

In the modern world, Video Conferencing Systems (VCS) have become an essential tool for remote communication. These systems enable individuals and teams to collaborate and hold meetings from a distance, reducing travel costs and energy consumption. However, as the use of VCS continues to grow, there is an increasing need to evaluate their security, privacy, data economy and sustainability aspects. In particular, there have been reported several security and privacy issues with regard to the use of VCS (e.g. [1] and [2]) which raises concerns in the field of multimedia security. Inherently, there are privacy issues with all VCS, as facial videos and speech audios are involved, which are considered as biometric data requiring special protection. Furthermore, by analysis of the activity timelines of participants, behavioral patterns such as absence from the conference, movements, chatting behavior can be derived. These facts suggest that security analysis, especially with regard to privacy, is an important requirement within the field of multimedia security.

Due to the closed-source nature of most commercial VCS, security and privacy analysis approaches are somewhat limited to binary code analysis or behavioral analysis. e.g. by analyzing network traffic during live VC-sessions. Although VCS communication nowadays is almost 100% endpoint encrypted, but network traffic analysis can still reveal meta information about information flows, including their volumes, temporary behavior, server endpoints, streams, locations and more. In addition to deriving privacy related findings such as server operators and locations, there is also the potential to infer aspects of data economy, sustainability and reliability by these multimedia data. The network traffic analysis approach of VCS also brings along privacy issues in itself, as it requires real persons to actually use the VCS under investigation, exposing their aforementioned sensitive traits to potentially untrusted VCS service providers.

A. Research Gap

These previously mentioned aspects leads to the need of a new approach to conceptually generate virtual video conferencing (VC-) sessions by injecting media data into VC clients to create reproducible virtual sessions for a comparable analysis of different VCS under the same data. Current concepts for the analysis of VCS (for example [1]) did not provide an environment with automated, simulated and virtual network traffic (video, audio, and text) to evaluate the systems. We assume that for a systematically forensic analysis and comparison of multiple VCS a novel concept is needed which provides simulated, stable and scripted network traffic data based on virtual VC-sessions without privacy concerns.

B. Contribution

In this paper, we introduce a novel approach for the analysis of VCS using publicly available videos and pre-programmed scripts to simulate other VCS related activities like chatting, screen sharing and so on to generate network traffic. This approach allows the forensic analysis of encrypted data without the need for decryption and enables the evaluation of VCS without compromising user privacy. The use of virtual data instead of real biometric data protects the privacy of individuals and allows researchers to test and improve their algorithms and systems in a controlled and ethical manner and with the possibility of reproduction. The approach can provide a large and diverse dataset for studying the behavior of Video Conferencing Systems, enabling a more comprehensive understanding of these systems in real-world scenarios. Furthermore, the use of virtual data can lead to more accurate and reliable results compared to using noisy and unpredictable biometric data. Our methodology for generation of virtual VC-sessions to analyze VCS involves five steps:

1) Definition of user activities,
2) Data requirements and collection of data,
3) Automation of virtual VC-session,
4) Capture network data from virtual VC-session and
5) Forensic Analysis.

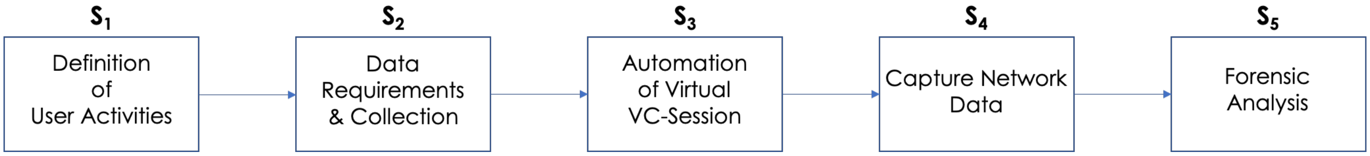
S1
Definition of User Activities
S2
Data Requirements & Collection
S3
Automation of Virtual VC-Session
S4
Capture Network Data
S5
Forensic Analysis

Fig. 1. Pipeline of our novel methodology

While our approach is still in its initial stages, we believe that it has the potential to offer a novel framework for the forensic analysis of VCS in future work.

C. Structure

The work is structured as follows: In Section II, a selected State-of-the-Art of VCS analysis is presented without the claim to completeness. In Section III the novel methodology is introduced. Section IV concludes the paper with a summary, early results and future work.

II. Selected State-of-the-Art

Forensic analysis of Video Conferencing Systems (VCS) has gained attention due to the increased use of these systems during the last years to save health, energy, costs and time. With the widespread relocation of the workplace to a private environment, security- and privacy-related aspects of Video Conferencing Systems have become more important.

Early works primarily addressed *Microsoft's Skype*™ software as widely used VCS during the last decade. Relevant papers address the following topics:

- Physical memory analysis to reconstruct user activities [3],
- Identification of *Skype*™ packets in network traffic [4],
- NAND and RAM analysis of *Skype*™ using an emulator [5] and
- Forensic analysis of *Skype*™ behavior on hard drive image on *Windows 10*™ [6].

The authors of [7] have investigated the *Cisco WebEx VCS*™ application in 2021. A forensic analysis of memory, hard disk, and a recording of network traffic was performed.

In [1] Altschaffel et al. presented an approach of a forensic examination process based on heuristics and meta data analysis of VCS-related multimedia network streams. Based on the general definition of seven multimedia data streams (audio, video, screen-sharing, sharing of video, text, file-transfer and other spatial streams) 20 events could be identified that revealed sensitive user or activity related information.

III. Methodology to Generate Virtual Sessions for Analysis of VCS

Our methodology for generating virtual video conferencing (VC-) sessions to analyze different Video Conference Systems (VCS) consists of five basic steps (S_1 - S_5, see Fig. 1). In the initial step S_1 the user activities for the virtual VC-sessions are defined (see Section III-A). In the second step S_2 the video and audio data without privacy concerns is gathered. The requirements for this data are described detailed in Section III-B. During the third step S_3 the automation of a virtual VC-session based on a screenplay (from S_1) is implemented (see Section III-C). In S_4 the network traffic produced by the virtual VC-session is recorded and post-processed (Section III-D). The last step S_5 features the potential analysis of the VCS with the recorded virtual VC-session (Section III-E).

A. Definition of User Activities (S_1)

In the first step of the methodology for analyzing VCS, we define user activities A_n (in this work, n denotes an index number for each individual activity A_n, event E_n or user U_n) to trigger VCS relevant events E_n (based on [1]) in order to ensure a consistent and reproducible evaluation process. This includes creating a screenplay, which specifies different activities A_n that users would carry out during the VC-session. A_n are based on the functionalities present in the particular VCS that shall be analyzed in S_5 with the intention of identifying specific events E_n of VCS. The screenplay includes details such as the number of users U_n who will participate in the session and what activities A_n they will perform, for example: join and leave the session, start sharing their virtual camera and audio and so on. By defining activities A_n in advance, we can control user behavior and ensure consistency across different simulations, allowing for more accurate comparison of different VCS. This step is critical for eliminating variations that could potentially influence an evaluation or a comparison of different VCS and enables a more reliable analysis (in S_5). The output result of this step is a screenplay that details the activities A_n for users U_n in the virtual VC-session.

B. Data Requirements and Collection of Data (S_2)

After defining the activities A_n for users U_n in S_1, the next step S_2 is to collect the media data required for the analysis of various VCS. Our primary goal in this step is to collect video and audio data without privacy concerns. To ensure this, we collect our data from publicly available sources such as news video platforms, which typically obtain consent from the individuals being recorded before making the footage available for public viewing. The collected data reflects the types of content and communication styles commonly encountered in real-world video conferencing scenarios. This helps us to create more accurate and relevant virtual VC-sessions. Before the collected data is used for analysis, it should be pre-processed to comply with defined requirements and standards. This may involve tasks such as formatting the data in a specific way (e.g. converting video to a particular file format or resizing of video media to a defined resolution to simulate a web cam), removing any irrelevant or extraneous information such as in-image text, or applying various filters or transformations to the data. The output result of this step is the collection of publicly available video and audio data for use in the subsequent steps.

C. Automation of Virtual VC-Session (S_3)

Step S_3 of our methodology is the automation of the VC-session. Therefore a simulation script is implemented that takes different media like audio, video and text data as input and carries out activities A_n for users U_n as defined in S_1 in a chronological order without the need for manual intervention. As a result this step reduces the workload of manually carrying out activities A_n on different clients and allows a reproducible data recording. What programming language is used for scripting the activities A_n of users U_n from S_1 with the collected data from S_2 should be decided by individual preferences (we used *Python 3.10*).

D. Capture Network Data from Virtual Session (S_4)

In Step S_4, network traffic data is captured from VCS of defined user activities A_n from S_1 with collected multimedia data from S_2, automated with the simulation script from S_3. For capturing network traffic data, we utilize a Switch SW that mirrors the traffic of all *Ethernet* connected systems (from users U_n) to a Data Collector DC (see Fig. 2). Thus, DC captures the network data. It is noteworthy that in our work we can only gather network traffic data for the client-side, since we usually do not have access to the server-side (*VCS-SV*) for most commercial VCS. This approach guarantees that the data we collect accurately reflects real-world VCS scenarios. When the data is captured, it has to be post-processed to ensure that it contains only relevant data for the subsequent and use case specific forensic analysis in S_5 (e.g. filtering out network traffic communication based on the IP addresses of client and server along with filtering only a single protocol for analysis like *TCP* (Transmission Control Protocol) or *UDP* (User Data Protocol). Thus, S_4 provides a post-processed dataset of selected network traffic of a virtual session for a specific VCS scenario, with controlled and consistent user behavior.

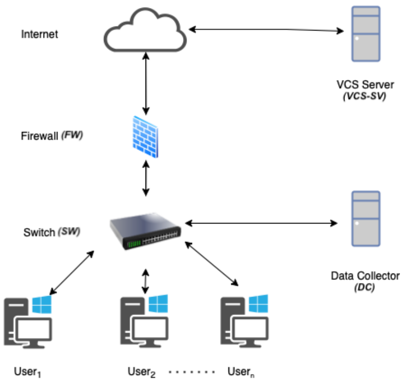

Fig. 2. Setup to capture network data of virtual VC-session S_4

E. Forensic Analysis (S_5)

The final step S_5 of the methodology involves the forensic analysis of the data provided from S_4. The analysis can focus on various use case specific aspects, such as privacy, reliability, user activity tracking, sustainability and more. The specific aspects chosen for analysis will depend on the goals and objectives of a potential study. To perform the forensic analysis, various statistical computational techniques such as machine or deep learning based approaches can be taken into consideration. For an initial validation of the concept, we have carried out an exemplary user activity tracking for S_5 (see Section IV-B).

IV. Conclusion

A. Summary

In this paper we introduce a novel methodology for the generation of virtual session for VCS to provide a forensic analysis of VC-sessions with reliable, comparable and reproducible data. The methodology involves four steps to create data for virtual session. In the fifth step of the methodology a forensic analysis is performed which can focus on various aspects such as privacy, reliability, sustainability, user activity tracking and so on.

B. Early Results and Future Work

For the initial proof-of-concept validation of the new methodology presented in this paper, we perform a first user activity analysis based on visualization of multimedia streams. During this initial exemplary analysis (step S_5), we identified user activities based on events from [1] like webcam on/off, muted/unmuted and screen sharing on an exemplary VCS. Thus, we can expect that our concept can provide reliable and reproducible virtual VC-session, which can be used for the analysis of VCS.

We will use the methodology to provide further virtual VC-session for a machine learning based analysis of multiple VC-Systems with the intention of user activity tracking.

References

[1] R. Altschaffel, J. Hielscher, S. Kiltz, and J. Dittmann, "Meta and media data stream forensics in the encrypted domain of video conferences," in *Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security*, 2021, pp. 23–33.

[2] MITRE-CVE-Program, "CVE-2022-36927," Available from MITRE, CVE-ID CVE-CVE-2022-36927, 2022. [Online]. Available: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2022-36927

[3] M. Simon and J. Slay, "Recovery of skype application activity data from physical memory," in *2010 International Conference on Availability, Reliability and Security*. IEEE, 2010, pp. 283–288.

[4] A. Azab, P. Watters, and R. Layton, "Characterising network traffic for skype forensics," in *2012 Third cybercrime and trustworthy computing workshop*. IEEE, 2012, pp. 19–27.

[5] M. I. Al-Saleh and Y. A. Forihat, "Skype forensics in android devices," *International Journal of Computer Applications*, vol. 78, no. 7, 2013.

[6] A. Majeed and S. Saleem, "Forensic analysis of social media apps in windows 10," *NUST Journal of Engineering Sciences*, vol. 10, no. 1, pp. 37–45, 2017.

[7] Z. Khalid, F. Iqbal, F. Kamoun, M. Hussain, and L. A. Khan, "Forensic analysis of the cisco webex application," in *2021 5th Cyber Security in Networking Conference (CSNet)*. IEEE, 2021, pp. 90–97.