Surveillance and Datenschutz in Virtual Environments

Sabine Cikic[1], Fritz Lehmann-Grube[1], and Jan Sablatnig[2]

[1] Technische Universität Berlin, Center for Multimedia in Education and Research (MuLF), Sekr. MA 7-2, Str. des 17. Juni 136, 10623 Berlin, Germany
{cikic,lehmannf}@math.tu-berlin.de

[2] Technische Universität Berlin, Institute of Mathematics, Sekr. MA 7-2, Str. des 17. Juni 136, 10623 Berlin, Germany
jon@math.tu-berlin.de

Abstract. Virtual environments are becoming more and more accepted, and part of the everyday online experience for many users. This offers new potential for both surveillance and data mining. Some of the techniques used are discussed in this paper.

However, such activities may in many countries conflict with the legal framework in place, for example with the German Federal Data Protection Act (Datenschutzgesetz). This point is illustrated by means of comparisons with real-world collection of personal data scenarios such as telephone tapping or video surveillance.

Keywords: virtual surveillance, virtual environment, privacy.

1 Introduction

With the widespread success and increased user numbers of virtual environments such as *World of Warcraft* and *Second Life* come new challenges and opportunities for law enforcement and intelligence services of all kinds. There are many reasons why one would want to observe and monitor users and their avatars (their digital representations while active in the virtual world).

Firstly, virtual environments may be used as a hitherto unmonitored base for inconspicuous meetings between real criminals [1]. Secondly, real crimes are commited that are related to virtual environments, either because the people involved met in a virtual environment, or because of actions taken by these people within the virtual environment [2]. Finally, observation of an individual's virtual actions may be part of a real-life observation of the individual, perhaps because of a real-life crime investigation or because private individuals (such as employer or spouse) have an interest in finding out about the person's whereabouts and contacts.

The collection of such data may, however, be at odds both with the TOS (Terms of Service) of the virtual environment and with the privacy laws of the country the subject is operating in (or that the virtual environment is operating in). This subject is still under debate, but possible conflicts will be illustrated and discussed, using the example of *Datenschutz* – the very strict privacy concept in force in Germany.

M. Sorell (Ed.): e-Forensics 2009, LNICST 8, pp. 212–219, 2009.

2 Surveillance

As virtual worlds consist entirely of data, it is possible to compile a very complete picture of the user's activities in these virtual worlds – if one has access to the data. Different means of gaining such access are available according to the privileges of the monitoring party.

2.1 Surveillance by ISPs

If you have access to the entire datastream going to and from the monitored user's computer, it is possible to record all data going to and from the virtual environment client. This record can later be used to reconstruct the user's entire virtual environment experience on a different computer. This then represents full surveillance of a single user within the virtual world.

There are several drawbacks to this approach. First of all, analysing the data and creating a log of events demands a large amount of PC processing time (for visualising the data), and a human expert would be required as well to interpret the on-screen data. This operator would have to watch the playback video and translate the images into a condensed description (such as "The suspect enters the house at 9:12pm").

Hence this approach is extremely inefficient, and unfeasible for mass surveillance of a large number of users. Similarly, this approach cannot be used to automatically find users that are engaging in any of a selection of activities on a watch list.

Additionally, the co-operation of the ISP (Internet service provider) must be obtained or the user's computer must otherwise be tampered with, which is an invasive technique liable to be regulated as tightly as telephone tapping.

Finally, it is theoretically possible to encrypt the transmitted data and thus make it difficult or even impossible to allow reproduction just by observing the data sent on the Internet. As yet, however, no virtual environment that offers encryption is freely available.

Surveillance by Internet tap can be neither detected nor avoided by the targeted user.

2.2 Surveillance by Administrators

A more promising point of access for observing a virtual environment is generally the provider's platform. The administrators of the virtual environment are typically able to easily collect and mine the entirety of data in the virtual environment. This also allows for complete surveillance, both of a single user and, in fact, of the entire virtual populace.

The advantage of this scheme is that it can be very efficiently implemented, as the data is there already and only needs to be filtered. It is also possible to, for example, filter for semantic information such as "Avatar enters house", rather than raw data such as "Avatar moves to position (x,y)". This makes mass surveillance easy. In fact, most providers of virtual environments already employ a large number of automatic watchdogs to alert their administrators to actions that the provider is aiming to prohibit in their virtual environment [3,4].

Co-operation with the provider of the virtual environment is necessary in implementing this technique. This can often be difficult to attain as the provider may reside in a different country and therefore lie outside of the jurisdiction of a surveillance order on a particular user or a user group.

It is also worth noting that third-party services such as voice-chat can be used in conjunction with virtual environments (although some offer built-in voice functionality). This information may then be missing from the final surveillance results.

Surveillance by administrators can be neither detected nor avoided by the targeted user. Also, this is the only type of surveillance that might yield answers to questions on *past* occurrences, such as "Was this user logged in last Friday?".

2.3 Surveillance in Programmable Worlds

A vast amount of data about other users can be gathered by simply being logged into the system as a regular user. This is especially the case where there is a programming interface that allows users to run code on the servers, as in *Second Life* or in MUDs (Multi-User Dungeons).

This type of observation requires a standard end-user computer with an account in the virtual environment, along with a good understanding of the programming interface that is used to create the virtual apparatus in question. Alternatively, many of the required virtual objects (such as listening devices) are already freely available from third parties.

Mobile Listening Device. It is easy to create a listening device that simply records all text chat. It is usually not difficult to make this device nearly undetectable – either very small or entirely transparent. The device can either be smuggled into the user's inventory or can be self-propelled and auto-following so that it will continuously watch an avatar while always staying at just the maximum "hearing" range from it.

These devices exist within the virtual world, and this is where they aggregate, filter, and collect their data. The devices available in *Second Life* are able to transmit their data to any computer outside of the *Second Life Grid* for further processing [5].

Where voice-chat is enabled within the world, this voice chat can likewise be captured.

Countermeasures exist for such devices, such as *Second Life*'s skyboxes (special rooms that control access by specific permission), chatbug scanners (that are trying to find listening devices nearby), or encryption tools providing simple in-world chat encryption [6]. As the virtual environments themselves are not meant to provide privacy, however, add-ons such as these are clumsy and not much used.

Stationary Listening Device. Listening devices may also be employed to watch a certain area and transcribe all chat there. Such devices may be deployed covertly or even openly.

These devices can be countered in a number of ways, especially if their presence is known, e. g. by refraining from substantive discussion while observation is in place, by teleporting through the area or moving in an otherwise unexpected way to avoid detection.

Camera. Within a virtual environment, a camera is very similar to a listening device, except that in addition to chat it can also store positions, movements, and the actions of nearby avatars.

The data thus collected can later be analysed and visualised to be displayed as a conventional video of what the camera saw.

Gridscanner. A single robot may not be able to easily travel through the entire virtual environment, as the process may take too long if the environment is very large. If the robot is able to clone itself, however, it is possible to scan the entire world and gather information, e. g. the positions of all avatars seen by any of the clones.

The gathered information may be collected to create online profiles of certain avatars, or may be used to find specific avatars and then use one of the techniques described above (e. g. a listening device) on them.

This scanning approach is a form of mass surveillance, and as such cannot be easily prevented by users. It is possible to avoid detection by staying in private rooms, especially if the scanning schedule is generally known.

2.4 Surveillance in Gameworlds

Even if a user has no special rights and no programming interface to the virtual environment, he can still gather large amounts of data about other users.

All of the following techniques require a computer with an account in the virtual environment, as well as a "bot" – a program that can interface with the client on the computer to act autonomously in the virtual environment. Bots are widely available for most popular virtual environments. Finally, transmission interception is required to record the exchange of data between the virtual environment and the client on the monitoring PC. This datastream can then easily be scanned for the information of interest. Although this requires some understanding of the stream format in question, the relevant information on these formats is generally publicly known. The result of the stream analysis then yields the surveillance data. Some of the most effective surveillance techniques using this set-up are listed below.

Who-Scanning. Many virtual environments allow a `who`-like command which shows all users currently logged in, or sometimes just all users in the current zone. In these cases, it is very easy to build a complete list of all avatars present, the times they are on-line, and optionally their area preferences (if this information is given in the world's `who` command). Similar weaknesses include `finger`-type commands which may reveal additional information such as last login time.

The easiest way to gather this information is to have the bot log in every once in a while and issue the corresponding command. The returning data is easily intercepted and compiled.

If information is given only for the current zone, the bot will have to teleport through all zones for each cycle.

Cityscanning. Another way to find attendance lists in a virtual environment is to simply walk through the main congregation points of the world (the cities). By intercepting and logging all avatar names that come into view-range, it is possible to generate incomplete but still considerable amount of avatar attendance data.

Again, the bot will have to walk through each city, teleporting from one to the next.

Chatlog. Most virtual environments offer a conveniently large area of audibility for chat messages. Thus, an entire city's chat can be heard from maybe two or three positions. In addition, there are often world and trade channels that everybody can talk on and listen to, regardless of position.

Again, chat streams are among the easiest to isolate from incoming data. In conjunction with city-scanning, this can yield chatlogs of many but not all of the conversations going on in-game.

If a user wants to collect complete chatlogs, he needs to employ many bots to stay in specific positions in the cities and thus listen in on all conversations in their local area. On the other hand, he can use just the world- and trade-channels as an incomplete variation of the `who`-command.

Trailing. You can also attempt to follow an avatar in a virtual environment, just as a real person can be followed in the real world. This will likely be noticed sooner or later, but with intelligent use of hiding places, changes of costume and of avatars, teleporting, flying, and similar techniques undetected pursuit may also be possible for a while. Actual eavesdropping is then often quite easy, as because of the simple, range-based in-game-logic of most virtual environments, there are usually many places where one can be completely visually occluded but still be within audibility range of another avatar, especially when one of the avatars in question is inside a building. The success of such an endeavour depends both on the chat-audibility range in the virtual environment and the aptitude of the follower. The downside of this is that in general bots cannot be employed for this problem, instead a human controller is required, which makes such surveillance expensive.

It should be noted that many of the techniques described in the last two sections may violate the TOS of the virtual environment in question. In general though, the administration of the virtual world would not specifically scan for such occurences and so they will usually go undetected, with the possible exception of the massive botting involved in cityscanning.

3 Datenschutz

3.1 Origin and Application of Privacy Laws

The precedent for every modern regulation on the collection and flow of personal data is the concept of privacy formulated by the US lawyers Warren and Brandeis over a hundred years ago in 1890 [7]. With the rise of mass media and new technologies, new possibilities had come up to harm individuals by exposing them to the public eye. So Warren and Brandeis introduced the term "privacy", stating "That the individual shall have full protection in person and in property (...)".

In the 1970s, West Germany experienced a surge of terrorism, kidnapping and an principle ideological challenge to the democratic system, which was generally agreed to be the major crisis in the history of the state. Against this backdrop, attempts were made to strengthen democracy wherever possible, which lead to the first German privacy laws. The reasoning behind these laws was that since an atmosphere of surveillance tends to boost conformity of behaviour, the diversity of opinions is lessened by this surveillance. This effect is well-known and used in many totalitarian states (a literary example of this would be Orwell's *1984*), but to a democracy, it is harmful.

Datenschutz (pronounce "dartenshoots") literally translates as data protection, and not to data privacy. The reason for this diction was that these laws were intended to deal with the first forms of mass data collection and analysis that became possible in the 1970s with early computing systems. As such, the German law has very explicit rules as to when, by who and how any amount of data may be collected, if that data can in any way be linked back to individuals. Note that it is not specified what kind of data is targeted by the law, but the law was already applied to personal data (such as name, address, birthday), speech data (such as telephone conversations), and location data (such as gathered by video cameras on public spaces). Alongside a focus on violations of the privacy of an individual by *usage* of personal data, it also regulates *acquisition* and *processing* of the data itself. The law is based on the assumption that it can be harmful to the general democratic ethos if people feel intimidated by the recording of their behaviour.

The German law of Datenschutz is renowned as one of the strictest privacy laws in the world, and is often seen as a major barrier to information flow. To allow the gathering and processing of a collection of data, either a specific law, regulation or court order must be acquired, or each person whose data is to be collected has to state their consent, knowing also the exact type of data and by whom and how it is to be handled. If this is impossible, the data either may not be gathered at all or must be made anonymous before collation.

3.2 Application to Virtual Environments

The fundamental question on the legality of collecting data as detailed in section 2 is whether avatar-linked data should be regarded as user-linked. The provider of a virtual environment can usually access information on this connection, or can at least find it out by applying an IP trace. Yet the information is not common knowledge and so one could take the position that data collected within a virtual environment is not personal data and therefore need not be protected at all. On the other hand, the directive 95/46/EC of the European Parliament applies, which explicitely states, that "account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person" [8]. Following this, in-game behaviour must be regarded as user-linked.

Another question is whether the data that accrues within the virtual environment can even violate a user's privacy even if everybody knows who the user is. This may depend on the virtual environment in question. Blizzard, maker of the popular *World of Warcraft* game, for instance, has a privacy statement on their website [9] that discloses very specifically how personal data such as the user's address, age and payment history are handled, but this statement does not mention in-world data such as avatar level, log-on times or alliances. Such data is regarded as actively published, impersonal artefact and in no way concerning the player behind. Linden Lab on the other hand claims, that a *Second Life* avatar should be an extension of the personality of a real person, and encourage users to express themselves as easy as in reality. So here the user's privacy can be harmed by surveilling his avatar.

Either way, some of the above practices have very close real-world analogs which can be used to decide whether the practices are probably legal.

ISP Surveillance. This is very much akin to telephone tapping and can in fact be used for that purpose in the case of VoIP. Well-known rules and regulations apply to this practise.

Virtual Cameras. In real life, there are strict rules in Germany as to how long data from a surveillance camera may be retained. This allows this data to be used in case of a criminal incident, but no general analysis and aggregation of the data is allowed.

Gridscanning, Cityscanning. In the real world, the current position of most of the population can be analysed from the cellphone-cells where they are currently registered. In Germany, this kind of data on a specific individual can only be collected with the individual's consent or on a court order.

Listening to Voice-chat. This is a direct analog to telephone tapping and as such is illegal without an explicit court order.

Chatlog. Both real-world mail and telephone conversations are specially protected by law in Germany and, by common extension, so is e-mail and other types of private exchanges of opinion such as instant messenger logs, IRC logs, or chatlogs.

4 Conclusion

Users in virtual environments are generally not highly aware that they may be observed within these environments. This may have to do with the fact that they are regularly monitored by administrators and thus are used to a certain level of surveillance. On the other hand, users usually have a high feeling of security from surveillance. There may be several reasons for this, such as the belief that it is either too expensive or too uninteresting to continually monitor them, as long as they don't do anything conspicuous. Also, virtual environments' principal feature of slipping into a made-up avatar character tends to disassociate users from their online-persona and thus the need to protect their privacy, as well.

Finally, the similarities of virtual environments with the real world tempts users to assume that if they cannot see anybody, they cannot be overheard.

However, while a real-world listening device still costs a few dollars, a *Second Life* listening device costs mere cents and it can clone itself to keep watch on thousands of avatars at the same time. We have illustrated a few particularly cost-effective methods for gathering information on avatars from a virtual environment in this paper.

But any organisation collecting such data must consider the legality of its actions. In Germany, most of the methods described above would *probably* be deemed illegal, especially when executed anonymously. Note however, that the situation becomes legally much more involved when several countries are involved, e. g. if an Australian logs onto an US-American gameserver to gather information on an Italian user.

References

1. Toavs, D.: Emerging Media: Its Effect on Organizations. Talk at DNI Open Source Conference (2008), `http://blog.wired.com/defense/files/OSC-TOAVS.ppt`
2. Chinese gamer sentenced to life, `http://news.bbc.co.uk/1/hi/technology/4072704.stm`
3. MDY v. Blizzard, `http://virtuallyblind.com/category/active-lawsuits/mdy-v-blizzard/`
4. Inside Club Penguin and its Child Safety Program (PlayNoEvil Game Security News & Analysis), `http://playnoevil.com/serendipity/index.php?/archives/1461-FEATURE-ARTICLE-Inside-Club-Penguin-and-its-Child-Safety-Program.html`
5. Dodds, C.: Avatars and the Invisible Omniscience: The panoptical model within virtual worlds. RMIT University (2007), `http://iconinc.com.au/christo/C.Dodds_Exegesis.pdf`
6. Chevalier Encryption HUD, `http://www.xstreetsl.com/modules.php?name=Marketplace&file=item&ItemID=385445&affiliate=10824e9e2587b0d00069c4efba3212`
7. Warren, S.D., Brandeis, L.D.: The Right to Privacy (1890), `http://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy_brand_warr2.html`
8. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML`
9. Blizzard Privacy Policy, `http://www.worldofwarcraft.com/legal/`