Forensic Content Detection through Power Consumption

Ulrich Greveler, Benjamin Justus, Dennis Loehr
Computer Security Lab, Münster University of Applied Sciences
D-48565 Steinfurt, Germany
Email: {greveler|benjamin.justus|loehr}@fh-muenster.de

***Abstract*—Digital forensic investigators are facing challenging problems of finding clues and solving crimes involving digital data. Advanced metering devices (smart meters) are being installed throughout electric networks in Europe and in the United States. The high-resolution energy consumption data which are transmitted by some smart meters to the utility company allow identification and monitoring of equipment within consumers' homes. Our research shows that the analysis of the household's electricity usage profile at a $0.5s^{-1}$ sample rate permits identification of audiovisual content. By collecting a single household smart meter data over a period of one month, our forensic results also show that audiovisual content identification is not heavily obstructed by typical electric appliance activities in a household.**

Keywords. Smart Meters, Digital Forensics, Audiovisual Identification, Privacy, Security

I. Digital Forensics

Digital data are ever increasing in quantity and complexity due to its omnipresence in our digitalized modern society. Digital forensic investigators are thus facing ever challenging problems of finding clues and solving crimes involving digital data. The number of forensic cases is ever mounting in the private sectors as well public sectors because of a growing awareness of better forensic techniques at all levels of law enforcements. Large forensic targets (data size on the scale of Terabytes or more) are nowdays accessible due to better data acquisition techniques, availability of distributed computing facilities, and the state-of-the-art forensic techniques.

A. Smart Meter Data

A smart meter is an electrical meter that records consumption of electrical energy at intervals and has the capabilities of communicating between a central server of its recorded information. By 2020 [1], the smart metering devices are supposed to replace 80% of the existing conventional meters in the European Union. Consumers using a smart meter are able to view their detailed energy consumption data via a web-browser, through which they can see into details how energy at home is used, therefore providing possibilities for devising energy saving strategies in view of their energy consumption habits. The energy company can also use the smart meter data for the purposes of infrastructure planning, network optimization and load balance checking.

The new generation of smart meter is able to provide fine-grained (interval of 2 seconds or less) power consumption data. High resolution data give data forensic researchers chances of introducing new forensic techniques arising from statistics [2], [3], network forensics [4], and computer forensics [5]. These new techniques will inevitably raise new privacy concerns in the future smart-grid infrastructure. A smart meter equipped household may provide malicious parties chances of probing into not only private household electrical consumption activities [6], but the personal habit of an individual as well [7], [8].

B. Our Contribution

Depending on the granularity of measurement and the resolution of data, we show in this paper that it is possible to deduce which video content an individual has viewed in the course of a smart meter recording. We also provide statistical results of our investigation. These results show that a forensic software based on our approach can be built in order to detect for example viewing of infringed video content. Our prototypical forensic tool is algorithm-based, and provide live analysis for any continuous aggregation of smart meter data.

C. Paper Outline

The smart meter hardware and the television hardware that are used in the test are described in section II. Section III contains detailed description of the forensic algorithms. The test results are presented in section IV. The related work and the conclusion are presented in section V and section VI respectively.

II. Hardware Background

A. Smart Meter Hardware

The smart meter hardware is acquired from a private company Discovergy GmbH (Heidelberg, Germany). The calibrated smart meter is installed in a typical private house in the region North Rhine-Westphalia, Germany. The new meter replaces the conventional digital meter which has been the default electric meter deployed by German public utility company. The smart meter hardware is manufactured by EasyMeter GmbH, Bielefeld (Electronic 3-phase meter Q3D-A1004 v3.03) and takes measurement at an interval of two seconds.

B. Television Hardware

During the forensic investigation, our experiments were carried out on normal household television sets and while other appliances were operational. During the film forensic investigation for example, we have carried out extensive tests on a Panasonic LCD Model[1] which happened to be in the tested household. The power consumption difference of a frozen white picture to a frozen black picture for this particular model was measured to be about 70 watts.

III. TV/Film Detection

We sketch below components of the forensic algorithm.

A. Power Consumption Prediction Function

The power consumption prediction function provides power usage prediction for the playback of a particular multimedia chunk. The input of the function is the multimedia content, the output is power usage prediction as would displayed by a smart meter.

The first step of the investigation involves the determination of the value b_{min}, which is defined to be the minimum brightness value that maximizes TV power consumption. To do so, we first measure the power consumption for a series of pictures consisting of elementary shades. The additive RGB color palette with one byte (i. e. values 0–255) per red, green and blue portion is used. The sequence of pictures are then RGB 0-0-0, RGB 1-1-1, ..., RGB 255-255-255 that increase the brightness from black to white running over 254 shades of gray. Our observation shows that maximum power consumption (i.e. b_{min}) is reached with rather dark pictures (e. g., RGB 32-32-32). But this also depends on the television user settings. A typical b_{min} value for the tested LCD TVs lies in the range $\{26, \ldots, 58\}$.

The next step (shown in Figure 1) is to extract frames from the movie and determine the brightness of each frame. The mean value of the red, green and blue portion is calculated to be the frame brightness value b (or value b_i for a frame with index i). By assuming a linear function (suggested by the results of step one) we can then let the predicted power consumption m_i (for a frame with index i) to be at the TV set's maximum power consumption for all frames being brighter then (*RGB* b_{min}-b_{min}-b_{min}) and being equal to $(max - min)(b_{min} - b)$ for all frames with brightness $b < b_{min}$. To be more TV device independent we use a function with values from 0 (minimum power consumption) to 1 (maximum power consumption).

$$m_i := \begin{cases} 1 & \text{if } b_i > b_{min} \\ \frac{b_i}{b_{min}} & \text{otherwise} \end{cases}$$

As we obtained our experimental results with a smart meter operating on a two-seconds interval, we then calculate an average value of power consumption for a number of consecutive frames adding up to two seconds of a movie, e.g.

[1]Panasonic model number TX-L37S10E

$$n := 2 \text{ times (no. of frames per second)}$$

$$m_i := \begin{cases} 1 & \text{if } b_i > b_{min} \\ \frac{b_i}{b_{min}} & \text{otherwise} \end{cases} \qquad pp_k := \frac{1}{n} \sum_{i=nk}^{n(k+1)-1} m_i$$

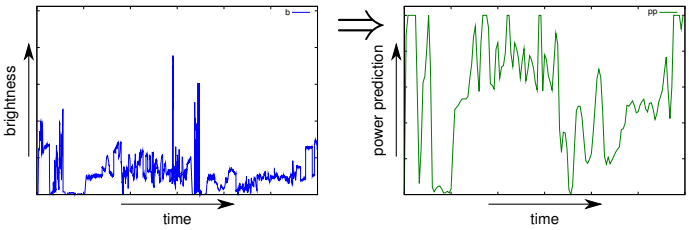

Fig. 1. Power prediction is computed on frame brightness values

50 frames for a movie with a typical 25 frames per second (fps) rate.

$$pp_k := \frac{1}{n} \sum_{i=nk}^{n(k+1)-1} m_i$$

Our derived power prediction function does then give a predicted power consumption value after 2s ($k = 1$), 4s ($k = 2$), 6s ($k = 3$), etc. This data can be correlated with any subsequent power profile data of the same length in order to search for the content.

B. Corridor Algorithm

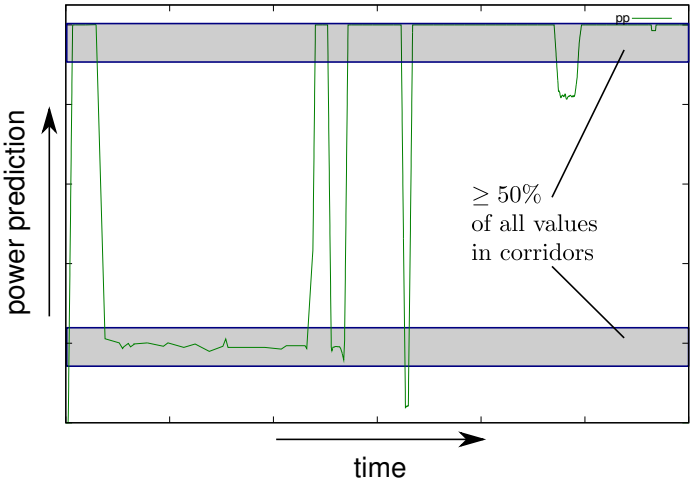

Fig. 2. Corridor algorithm discards chunk where more than half of values are found in distinct two corridors

During the experiments, we have noticed that the power consumption curve as observed by a smart meter oscillates in a normal household situation (without TV running) in a way that could lead to false positive identification of TV content. The reason for that is while searching for 5-minute chunks of movie files, if the chunk is for example showing a long dark scene, followed by a long bright scene (both scenes added exceed five minutes), it will correlate strongly with a power curve that reflects the switching-on of a simple electric

appliance (e. g. a light bulb). To make movie load signature more distinguishable, it is desirable to eliminate possible false matches reflecting this effect during the analysis stage. For that purpose, we have developed a *Corridor Algorithm*. If too many values of predicted or actual power consumption fall in one of two corridors, this movie-chunk will be discarded. Figure 2 shows a typical scenario, in which the green power curve is truncated within the corridors which are highlighted in gray.

C. Forensic Work-flow

This section describes the work-flow that are involved in the movie identification process. The Figure 3 illustrates all the steps involved. In fact, we have developed a software script that automates the forensic process as described in the figure.

The entire film is first divided into fixed intervals (e.g. 5 minutes) chunks, and the brightness of a frame in each chunk is calculated. The power prediction function (section III-A) is used to compute the correlation value for the chunk. The decision of discarding the chunk depends whether the correlation reaches the prescribed threshold value. The chunk is further processed with a b_{min}-optimization algorithm and the corridor algorithm if the correlation is greater than the threshold. A positive match is declared when the chunk survives the entire process.

It should be noted that the failure of the identification process (e.g. due to power disturbance or other reasons) on some video chunks does not pose a big threat when one has the entire movie at disposal. Since for a typical 90 minutes playback, we have $90/5 = 18$ blocks at disposal if we divide the movie into five minutes chunks.

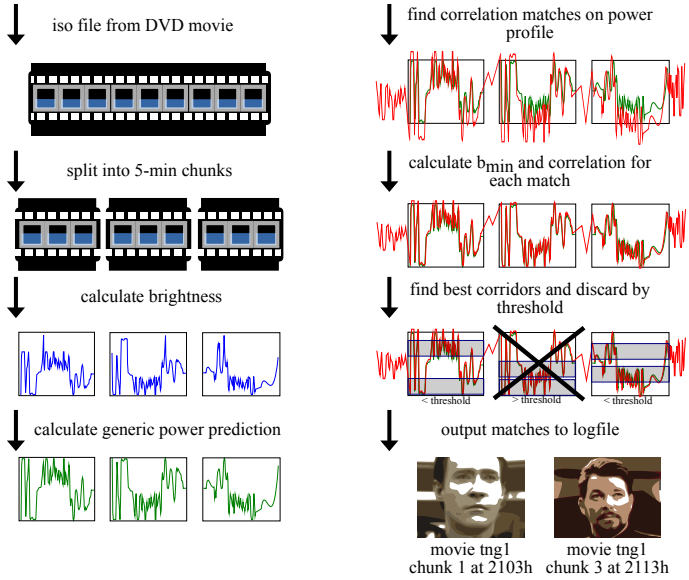

Fig. 3. Work-flow to detect chunks of a movie

We define a *positive identification of a content* as the event that at least two 5-minute-chunks of the same movie which coincide in their time offsets are found. The example in Fig. 3 shows such a positive ID with chunks 1 and 3 that are exactly 600s apart.

D. Remarks on Discoverable Material

We have noticed during our investigation that not all broadcasts content are detectable. In the case of daily news program[2], we fail the identification because lighting level of each content block consistently stays about the same level. This leads to an almost flat line in the power prediction curve for back-lighted LCD TVs and the fluctuation of power consumption therefore can not be detected. The detection also fails on some programs with many flash scenes exposures[3]. The failure can be attributed to the fact: there is a higher brightness level than typical b_{min} values.

IV. Test Results

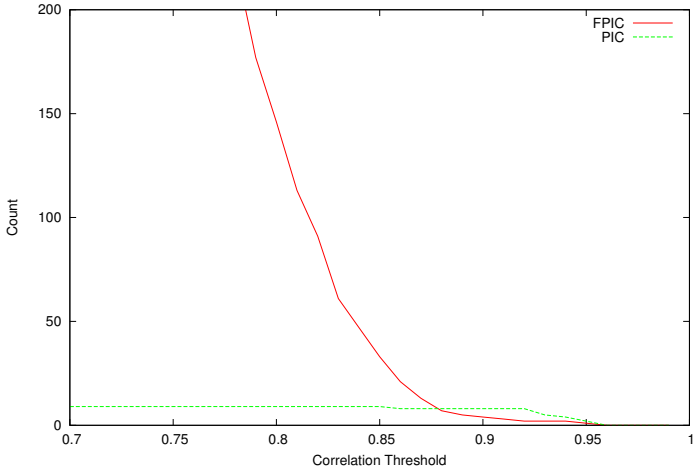

Fig. 4. Determination of correlation threshold values: counting false positive identified 5-min-chunks

In this section, we have collected some forensic statistical results. Our goal here is to test the robustness of the content identification algorithm as described in section III-C. The generic approach here is to run the forensic algorithms on various video content as described below.

Before starting the forensic work-flow, one requires the input of a threshold correlation value, on which the *corridor algorithm* is based to eliminate noisy movie chunks. To select a suitable correlation threshold value, we used a database consisting of 3414 movie chunks that spans over a playtime of 150 minutes. During the test, we randomly selected 10 movie chunks and played these selections twice. We recorded the number of positive identifications (PIC) and the number of false positive identifications (FPIC) and plotted them versus correlation values in Figure 4. The plot suggests a possible starting threshold value near 0.85.

Using the threshold values obtained above, we are now able to start the movie identification test. Table I shows a typical

[2]*ARD Tagesschau*: daily German news broadcast at 8 p.m.
[3]JAG. Director: Donald P. Bellisario, air date: 1995 - 2005

TABLE I
ENTIRE MOVIE TEST

Movie	Length	NoC[a]	FPCI[b]	PCI[c]
M1[d]	151 min	30	89	14
M2[e]	91 min	18	10	2
M3[f]	111 min	22	58	11
M4[g]	94 min	18	31	11
M1 2nd. Test	151 min	30	101	11

[a]number of 5 minute chunks
[b]false positive chunk identifications
[c]positiv chunk identifications
[d]Movie: *American Gangster* (2007). Director: Ridley Scott. Release Date: October, 19 2007
[e]Movie: *American Pie* (1999). Director: Paul Weitz. Release Date: July, 9 1999
[f]Movie: *Breach* (2007). Director: Billy Ray. Release Date: February, 16 2007
[g]Movie: *Transporter 3* (2008). Director: Olivier Megaton. Release Date: November, 26 2008

test scenario, in which four movies are played. The length of the movies and the number of chunks associated with it are recorded in column 2 and 3. Column 4 and 5 show the number of false positive identifications (FPIC) and the number of positive identifications (PIC) respectively. The test shows that we have two or more matches for all the movies played $[M1, ..., M4]$ with a correct time difference. The content with the smallest number of identified 5-min-chunks in this set is the movie *American Pie* (1999). For this movie we have identified only two chunks. This still leads to a positive content identification. Note Movie No. 1 was tested twice in order to get some experiences on the reproducibility of results.

A. Forensic Results

Forensic researchers are constantly required to reduce incident turnaround time, in order to better improve work efficiency. The detection of video content in smart metering data can be automated based on the algorithms described in section III. We have in fact developed prototype software for the purpose of automating the entire process.

To simulate a real life scenario, we have randomly played films out of a film database of 148 films that consists of 3414 5-minute chunks. We collected metering data over a period of one month in the test household. We then ran our forensic software tool on the data collected. The computation task was completed on a single server within 12 hours. Most content are able to be identified correctly as a result. Also we have found a few false positive content identifications within one month of arbitrary metering data. This can be attributed to the fact that a small number of chunks show a high correlation to activities of other electric appliances in the household. Table II lists false positive chunks which shows up more than 200 spots during the test. A possible improvement here is to discard all chunks that have a mean false positive count of three and more per day, we can then eliminate this false positive content identification and still able to detect every tested played movie.

TABLE II
FALSE POSITIVE RATE WITHIN ONE MONTH

Movie	Chunk No.	Number of found spots
Pride and Glory (2009)	4	2510
Dragon Ball (2009)	15	1565
Romeo + Juliet (1996)	7	1399
Elementarteilchen (2006)	17	1331
Rush Hour 3 (2007)	8	862
300 (2006)	15	597
Fatal Contact (2006)	16	426
Das Leben der Anderen (2006)	26	402
Pride and Glory (2009)	22	335
A Friend Is a Treasure (1981)	15	299
300 (2006)	21	299
The Three Investigators and the Secret of Skeleton Island (2007)	17	267
The Beach (2000)	19	240
Dio perdona... Io no! (1967)	11	236
Cyberflic (1997)	10	235

V. RELATED WORK

Multimedia forensic deals with tools and processes for gathering evidences in order to support investigators in their searches related to multimedia crimes. It is a vast research topic due to ever increasing amount of multimedia data on the web. Different aspects of multimedia forensics are discussed in the survey paper [9]. The common approaches in multimedia forensic investigation are: content fingerprinting and digital watermarking. In the context of movie protection, video fingerprinting can be used to provide in-house monitoring, identification of copyrighted content, and detection of pirated material [10]. Video fingerprint generation basically involves a two step process: key frames extraction and key frames characterization [11]. Digital watermarking requires modifying content prior to distribution. While remaining imperceptible to human observers, a watermark detector is able to retrieve embedded information in typical user cases. Video watermarking is possible via the location embedding mechanism [12].

Extensive researches have been done on techniques of non-intrusive load monitoring (NILM). Various NILM methods [13], [14] are introduced in order to glean into detailed energy consumption pattern in a household. Using these techniques, it turns out that a remarkable number of electric appliances in a private home can be identified by their load signatures with impressive accuracy. The same NILM techniques can be applied to analyze smart meter data in order to peek into household activities [15].

VI. CONCLUSION

We have demonstrated that content identification through fine-grained smart metering data is generally a feasible task. During the test, we had a database consisting of about 150 movies, from which the forensic results are deduced. In real situations, an investigator might want to search for all existing video material of one producer or all movies not yet released on DVD. In order to make our algorithms into a usable industry forensic software, one would have to have a much larger test database, and include many more test conditions

that exist in a normal household settings. Only then can we with statistical confidence assert the true values of various parameters in the algorithms (i.e.false positive ratio, threshold values).

Our forensic results nonetheless show that video content identification through power consumption profile analysis via a smart meter is not heavily obstructed by other electric appliance activities in a household. The high difference of power consumption for bright and dark scenes of 70 watts and more allow positive content identification via a television set that is running among many other running household appliances. Our prototypical forensic software could detect every movies (150 tested) under realistic circumstances with acceptable computation time.

Finally, one would like to point out the privacy implications of such forensic technology. It is conceivable that in the near future, an investigator is able to find out by scanning through all accessible smart meter data of one area who have viewed a particular video content in a given time period. One would inevitably sacrifice their private information such as daily TV viewing habits. These privacy related concerns might cause unease among potential smart meter consumers. In the mean time, the privacy implications of fine-grained metering data should motivate future research on privacy enhancing solutions on the smart grid infrastructure.

REFERENCES

[1] European Parliament, "Directive 2009/72/EC," Tech. Rep., 2009.

[2] U. Budhia, D. Kundur, and T. Zourntos, "Digital video steganalysis exploiting statistical visibility in the temporal domain," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 4, pp. 502–516, 2006.

[3] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," in *IEEE Workshop on Statistical Analysis in Computer Vision (in conjunction with CVPR)*, Madison, Wisconsin, 2003.

[4] S. Chen, K. Zeng, and P. Mohapatra, "Efficient data capturing for network forensics in cognitive radio networks," in *19th IEEE International Conference on Network Protocols*, 2011.

[5] G. G. R. III and V. Roussev, "Next-generation digital forensics," *Commun. ACM*, vol. 49, no. 2, pp. 76–80, 2006.

[6] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys 2010)*, Zurich, Switzerland, November 2010. [Online]. Available: http://www.cs.umass.edu/~kevinfu/papers/molina-markham-buildsys10.pdf

[7] "Researchers analyze smart meter data," http://www.spiegel.de/netzwelt/netzpolitik/0,1518,787629,00.html, September 2011.

[8] U. Greveler and B. Justus and D. Loehr, "background and experimental results on meter data," Muenster University of Applied Sciences, Tech. Rep. DAPRIM-2011, November 2011.

[9] S. Battiato, S. Emmanuel, A. Ulges, and M. Worring, "Second acm international workshop on multimedia in forensics, security and intelligence (mifor 2010)," in *ACM Multimedia*, 2010, pp. 1741–1742.

[10] M. Arnold, S. Baudry, P. Baum, X.-M. Chen, B. Chupeau, O. Courtay, G. Doërr, U. Gries, F. Lefèbvre, M. Morvan, A. Robert, C. Salmon-Legagneur, C. Vincent, and M. de Vito, "Multimedia security technologies for movie protection," in *Proceedings of the international conference on Multimedia*, ser. MM '10. New York, NY, USA: ACM, 2010, pp. 1515–1516. [Online]. Available: http://doi.acm.org/10.1145/1873951.1874267

[11] F. Lefèbvre, B. Chupeau, A. Massoudi, and E. Diehl, "Image and video fingerprinting: forensic applications," in *Media Forensics and Security*, 2009, p. 725405.

[12] D. Zou and J. A. Bloom, "H.264/avc substitution watermarking: a cavlc example," in *Media Forensics and Security*, 2009, p. 72540.

[13] H. Lam, G. Fung, and W. Lee, "A novel method to construct a taxonomy of electrical appliances based on load signatures," in *IEEE Transactions on Consumer Electronics*, 2007.

[14] A. Prudenzi, "A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recording at meter panel," in *IEEE Power Engineering Society Winter Meeting*, 2002.

[15] E. Quinn, *Privacy and New Energy Infrastructure*, Available: http://papers.ssrn.com/sol3/papers.cfm?abstract id=1370731, 2009.