Multimedia Content Identification Through Smart Meter Power Usage Profiles

Ulrich Greveler*, Peter Glösekötter‡, Benjamin Justus†, Dennis Loehr†
*Department of Communications & Environment, Rhein-Waal University of Applied Sciences, D-47475 Kamp-Linfort, Germany
ulrich.greveler@hochschule-rhein-waal.de
†Computer Security Lab, Münster University of Applied Sciences, D-48565 Steinfurt, Germany
{benjamin.justus, loehr}@fh-muenster.de
‡Department of Electrical and Computer Sciences, Münster University of Applied Sciences, D-48565 Steinfurt, Germany
peter.gloesekoetter@fh-muenster.de

***Abstract*—Advanced metering devices (smart meters) are being installed throughout electric networks in Germany (as well as in other parts of Europe and in the United States). Unfortunately, smart meters are able to become surveillance devices that monitor the behavior of the customers. This leads to unprecedented invasions of consumer privacy. The high-resolution energy consumption data which are transmitted to the utility company allow intrusive identification and monitoring of equipment within consumers' homes (e. g., TV set, refrigerator, toaster, and oven). Our research shows that the analysis of the household's electricity usage profile at a $0.5s^{-1}$ sample rate does reveal what channel the TV set in the household was displaying. It is also possible to identify (copyright-protected) audiovisual content in the power profile that is displayed on a CRT[1], a Plasma display TV or a LCD[2] television set with dynamic backlighting. Our test results indicate that a 5 minutes-chunk of consecutive viewing without major interference by other appliances is sufficient to identify the content.**

Our investigation also reveals that the data transmitted via the Internet by the smart meter are unsigned and unencrypted.

Our tests were performed on a sealed, operational smart meter used for electricity metering in a private home in North Rhine-Westphalia, Germany. Parameters for other television sets were obtained with an identical smart meter deployed in a university lab.

Keywords. Smart Meter, Data Privacy, Audiovisual Content, Smart Grid

I. Introduction

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. The installation of smart meters at private homes is planned in Germany, as well as in EU in the near future. By 2020, the smart metering devices are supposed to replace 80% of the existing conventional meters. Smart metering is believed to be a crucial factor for the future availability of supply, energy efficiency and renewable energy[3]. From a consumer perspective, smart metering offers potential benefits such as: consumers by using a smart meter are able to view their detailed energy consumption data via a web-browser. The visualization of these data lets consumer to 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. One also sees a trend towards IP-based communication as a common platform for smart meter applications[4].

Smart meter data contain consumer's personal information (see section IV). Also depending on the granularity of measurement and the resolution of data, we show in this paper that it is possible to deduce personal behavior of an individual in a private home. These behaviors include for example what TV channels, and which movies an individual has viewed in the course of a smart meter recording. In view of the concerns above, there are henceforth urgent calls for researchers to provide means of better protecting data transmitted by a smart meter.

II. Related Work

Even before the advent of smart meters, extensive researches have been done on techniques of non-intrusive load monitoring (NILM). Various NILM methods [12], [16] 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 [17]. More recently, the authors of [6] claimed that they were able to discern video contents from electromagnetic interference (EMI) signatures produced by different TV sets.

There have been privacy concerns over the deployment and usage of smart meters [2], [13], [18] in U.S. and Europe, precisely because they can inadvertently leak detailed information

[1]Cathode Ray Tube

[2]Liquid Crystal Display

[3]*80% Smart Meter Adoption By 2020 Through EU Mandate*: Yahoo Finance Report from Sep 29, 2011 8:10 AM

[4]*Its Official: The Future of the Smart Grid Is IP*: By Katie Fehrenbacher, Sep. 7, 2010, 7:57 AM, on gigaom.com

about household activities.There are currently two approaches of implementing privacy preserving smart meter data analysis. The first approach relies on masking the meter readings. The actual meter reading is adjoined by a masking value, in such a way that an adversary can not recover individual readings. Yet, the sum of the masking values across meters sums to zero. This technique is introduced [10], [11] to compute metering aggregation over a network. And most recently following this line, [4] developed a scheme that ensures the property of differential privacy. The second approach relies on homomorphic encryption. The metering aggregation using this approach is discussed in [7], and the algorithm within also allows detection of leakage in electricity distribution. Furthermore, in [5], [3] are introduced protocols to privately derive and prove the correctness of bills. Recently, a billing protocol based on Pedersen commitments and a plug-in privacy components is introduced in [9]. Finally, in [15] are introduced solution of embedding Trusted Platform Module (TPM) in the smart Meter to obtain signed tariff data.

III. Experimental Results

Our investigation aims to answer the following questions: (1) What are the possible ways of obtaining and evaluating data coming from a calibrated smart meter? (2) What can be deduced from smart meter data regarding a person's TV watching habit in a private home? The experiments mentioned in this paper took place from August to November, 2011.

A. Hardware Background

The tested smart meter had been acquired from the company Discovergy GmbH (Heidelberg, Germany) after signing a private household contract. This calibrated smart meter is installed in a typical private house in the region North Rhine-Westphalia, Germany. After the installation, the new meter replaces the conventional meter which is manufactured by the German public utility company RWE AG.

The Discovergy product is based on the smart meter model manufactured by EasyMeter GmbH, Bielefeld (Electronic 3-phase meter Q3D-A1004 v3.03). The smart meter takes measurement at an interval of two seconds. All data are transmitted to the servers hosted by Discovergy. The customers are then able to access these data via a web-browser. Discovergy[5] claims in its contract complete data encryption for each smart meter equipped household.

IV. Data Transmission

The transmission of smart meter data to the Discovergy-Server is done through the TCP/IP protocol. The meter is directly connected with a LAN/DSL router and receives its dynamic IP address via the DHCP protocol. Contrary to the company claim, the smart meter data are not encrypted. The energy consumption data are saved in a textfile format, while being transferred to the central servers. Figure 1 shows a snapshot of a typical data transmission. The unencrypted data can be easily hacked out.

[5]www.discovergy.com

```
POST /api/w.html HTTP/1.1
Content-Type: application/x-www-form-urlencoded
Host:85.214.93.99
Content-Length:851

version=0.9&identity=[redacted]&msg=228601&values=[
{"meterdata":"00000285.9823514*kWh","tickdelta":"00000285.9822239*kWh","seconds":"399511319.61"},
{"meterdata":"00000285.9824793*kWh","tickdelta":"00000285.9823514*kWh","seconds":"399511321.61"},
{"meterdata":"00000285.9826075*kWh","tickdelta":"00000285.9824793*kWh","seconds":"399511323.61"},
{"meterdata":"00000285.9827358*kWh","tickdelta":"00000285.9826075*kWh","seconds":"399511325.62"},
{"meterdata":"00000285.9828636*kWh","tickdelta":"00000285.9827358*kWh","seconds":"399511327.62"},
{"meterdata":"00000285.9829915*kWh","tickdelta":"00000285.9828636*kWh","seconds":"399511329.62"},
{"meterdata":"00000285.9831196*kWh","tickdelta":"00000285.9829915*kWh","seconds":"399511331.62"},
{"meterdata":"00000285.9832476*kWh","tickdelta":"00000285.9831196*kWh","seconds":"399511333.62"}]
&now=399511335.65
```

Fig. 1. Captured communication between smart meter and server

In addition, none of the data are signed. The identity (highlighted in black in Figure 1) of any smart meter is immediately revealed when the data are being transmitted to the central servers and could be used by an attacker to send different power consumption data to the server.

A. Resolution of data presented to the customer

Discovergy offers a web browser based view on the power consumption profile. A java-script based application requests the data from the Discovergy server and offers the visualization of the profile[6]. A typical profile example can be seen in Figure 2.

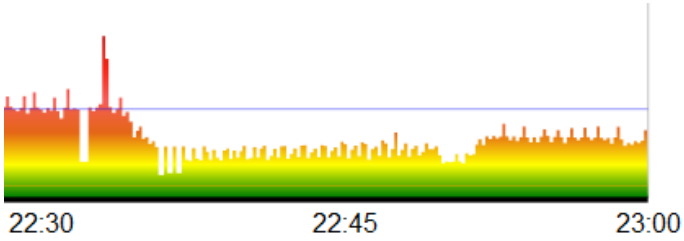

Fig. 2. Power profile visualized by Discovergy

An analysis of the script source code shows that the customer does not see the full resolution of the data (sampling rate at $0.5s^{-1}$). The data are consolidated by skipping absolute values (thus the arithmetic mean of several values is displayed). Moreover, a software bug regarding the time stamp parsing algorithm results wrong peaks and even negative peaks display. Such data would contradict the fact that the tested meter only submits monotonically increasing values (see captured communication in Figure 1). By correcting the software bug and downloading available meter data from the Discovergy server with a self-developed script we were able to visualize the complete data at various resolutions. Figure 3 depicts a small interval of data contained in Figure 2 (in the sub-timeframe 22.35h-22.50h).

B. Large Device Identification

We could verify the claims of other researchers [8], [13], [14]: electric appliances in a private home can be identified

[6]Note that this description reflects the state of the customer information portal *www.discovergy.com* during the test period (Aug. – Oct. 2011). It might be different at a later stage.

by their load signatures. In particular, we could identify the following household appliances: refrigerator, electric kettle, flow heater, light bulbs, energy-efficient lamps, bean-to-cup coffee machine, cooker hood, microwave oven, electric kitchen stove, washing machine, dishwasher, and the television set.

V. TV/Film Detection

A. Television Hardware

The first part of tests were performed on an home LCD television set in a household where the operational smart meter was installed. Liquid Crystal Display televisions use the display technology to produce colored images. Since the total amount of visible brightness of a picture is a combination of the backlighting and LCD shuttering, a technology dubbed *dynamic backlighting* is applied on modern LCD TVs to improve the contrast ratio[19]. While the shutters produce a contrast ratio of 1000:1, dynamic backlighting enhances this ratio up to 30000:1. The LCD TV power consumption is mainly influenced by the backlighting activities [1].

The experiment results presented in the following sections were obtained by using the household's Panasonic LCD television set[7]. Section V-H contains comparison results which use other TV models. The power consumption difference of a frozen white picture to a frozen black picture for this particular television was measured to be about 70 watts.

B. Power Consumption Prediction Function

The core of our content identification program is the power consumption prediction function. We explain below in details the construction of the function. The input of the function is the multimedia content, the output is power usage prediction as would displayed by a smart meter.

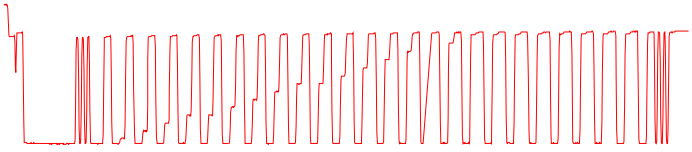

Fig. 3. Determination of b_{min}

The first step is to measure the power consumption for a series of pictures consisting of elementary shades. We use the additive RGB color notation with one byte (i. e. values 0–255) per red, green and blue portion. 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 is reached with rather dark pictures (e. g., RGB 32-32-32). But this also depends on the television user settings. For the rest of the paper, we denote this value by b_{min} which is the minimum brightness value that maximizes TV power consumption. A typical b_{min} value for the tested LCD TVs lies in the range $\{26, \ldots, 58\}$.

[7]Panasonic model number TX-L37S10E

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

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

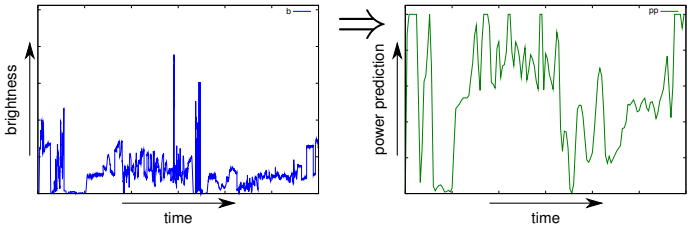

Fig. 4. Power prediction is computed on frame brightness values

Figure 3 shows one of the test runs we had performed in order to determine the value b_{min}. A sequence of pictures was shown: *black-white* (3 times) as a trailer to find the signal, then *black-(RGB-2-2-2)-white-black-(RGB-4-4-4)-white-black-(RGB-6-6-6)...* to see the increasing power consumption. One can then count the number of peaks until the gray picture (here: RGB-38-38-38) reaches maximum power consumption, i.e. becomes indistinguishable from white with regard to the power profile. At a later stage we did not need to run these tests anymore since we developed a script that performs content identification by automatic parameter detection.

The next step (shown in Figure 4) 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. 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

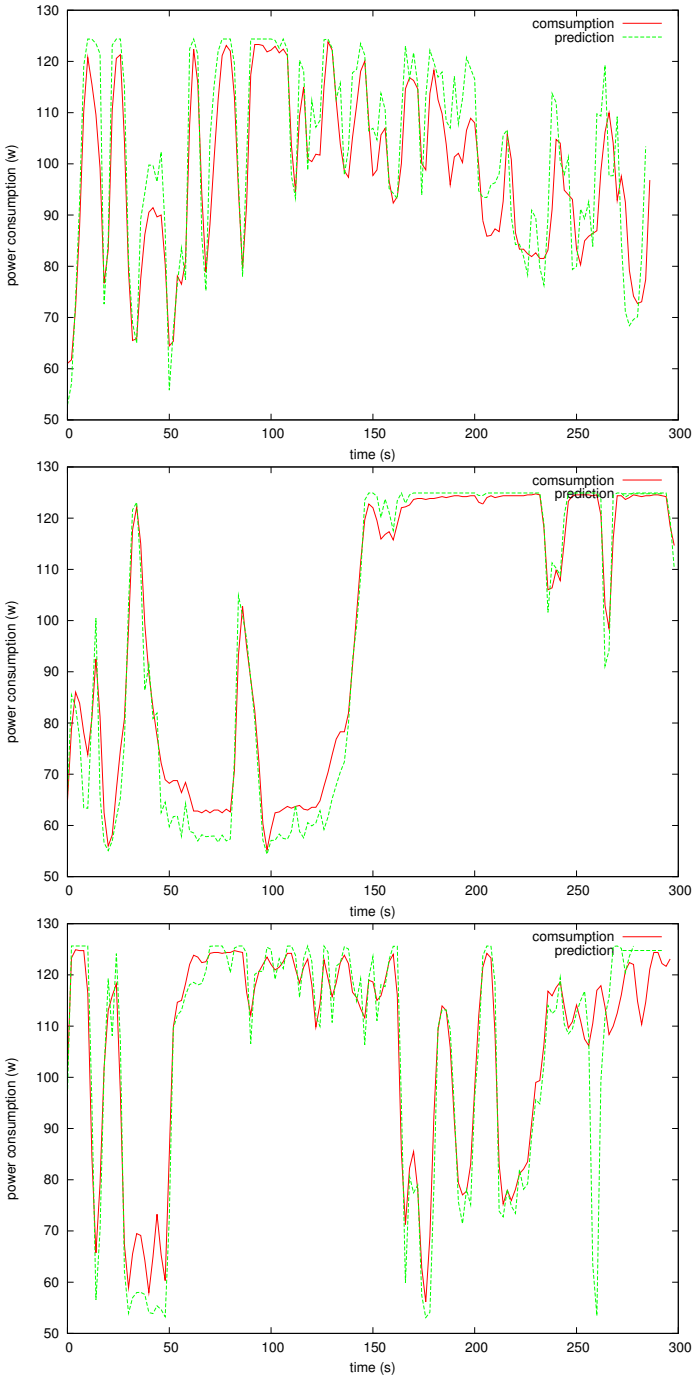

Fig. 5. power prediction vs. consumption: first 5 minutes of the movie Star Trek 11 (top), of episode 1, Star Trek TNG season 1, of the movie Body of Lies (bottom)

($k = 2$), 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.

C. Preliminary Analysis

To test our prediction function, we did a preliminary run on some films. We extracted first 5 minutes of each movie file, and then compared the actual power consumption against values produced by the prediction function. The movies we used for the test are:

1. Movie *Star Trek* (2009). Directed by J. J. Abrams. Release date: May 8, 2009.

2. Star Trek episode *Encounter at Farpoint* (1987). Directed by Corey Allen. Original air date: September 28, 1987.

3. Movie *Body of Lies* (2008). Directed by Ridley Scott. WarnerBros. Pictures. Release date: October 5, 2008

The actual power consumption was measured using a sealed operational smart meter while the films are playing on the household television set. No major appliances were operating during the measurements, only lights and stand-by consumption were active.

Figure 5 contains the experimental results. The green dotted curve is the prediction, and the actual power consumption data is plotted in red. We also calculated the Pearson product-moment correlation coefficients between the actual and predicted power consumption data. The correlation for the three movie events are 0.94, 0.98 and 0.93 respectively.

D. Corridor Algorithm

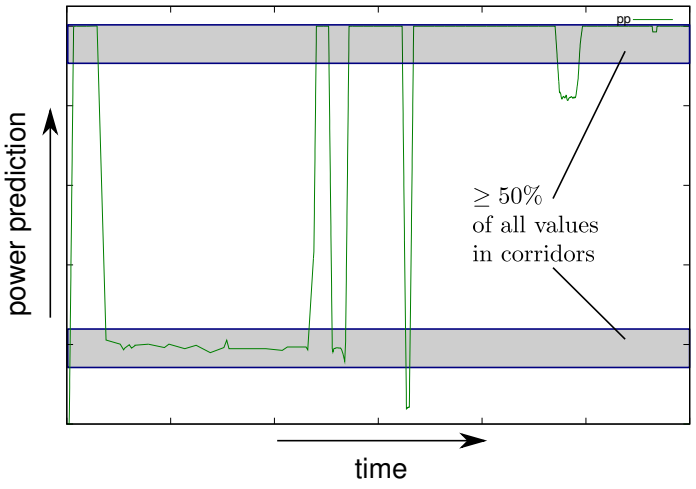

Fig. 6. Corridor algorithm discards chunk where more than half of values are found in distinct two corridors

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). So any curve jump phenomenon comes into the movie detection scenario (Picture 2 of Figure 5) showing a long bright scene could lead to false positive matching. 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 6 shows a typical scenario, in which the green power curve is truncated within the corridors which are highlighted in gray. The parameters for the decision (threshold, corridor heights) are derived in section V-I.

E. Automatic Detection of b_{min}

In order to identify a broad range of video material, we have developed a script to detect the optimal b_{min} values for each video content played. For each possible b_{min} value ($= 0 \ldots 255$), the correlation between actual consumption and power prediction is calculated. Figure 7 contains a comparison between actual power consumption curve with predictions supported by various b_{min} values.

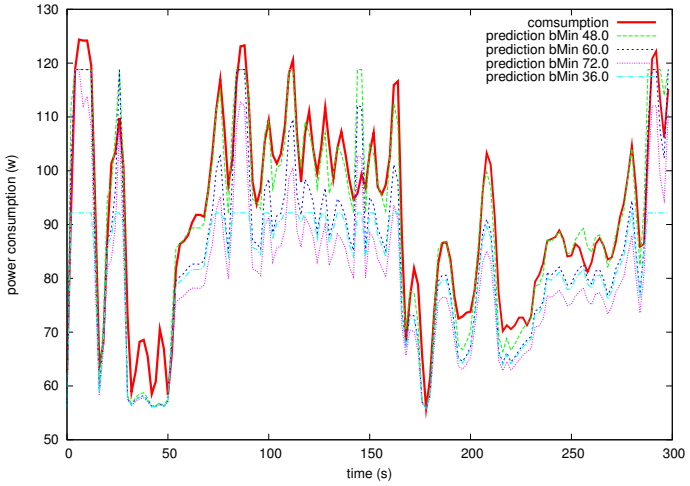

Fig. 7. Determination of b_{min}

F. Work-flow

This section describes the work-flow that are involved in the movie identification process. The Figure 8 illustrates all the steps involved. These steps are performed automatically by a software script we developed for the research being described in this paper and which could be regarded as a proof-of-concept for a forensic tool performing content identification on power consumption data.

The entire film is first divided into 5 minute chunks, and the brightness of each frame is calculated. The correlation value for the chunk is then computed using the predicted power values. The matches (correlation value is greater than 0.85 for a generic b_{min}-value) are further processed with a b_{min}-optimization algorithm and the corridor algorithm. The chunk is discarded if the threshold is reached. It should be noted that the identification process fails on some of the 5-minutes movie chunks due to either power disturbance or user interaction with the TV or the playing device. For a typical 90 minutes playback, we have $90/5 = 18$ blocks at disposal, so in a actual test there should be a good chance that at least two or three of these chunks *survive* other appliances' activities and can be found in the power curve matching.

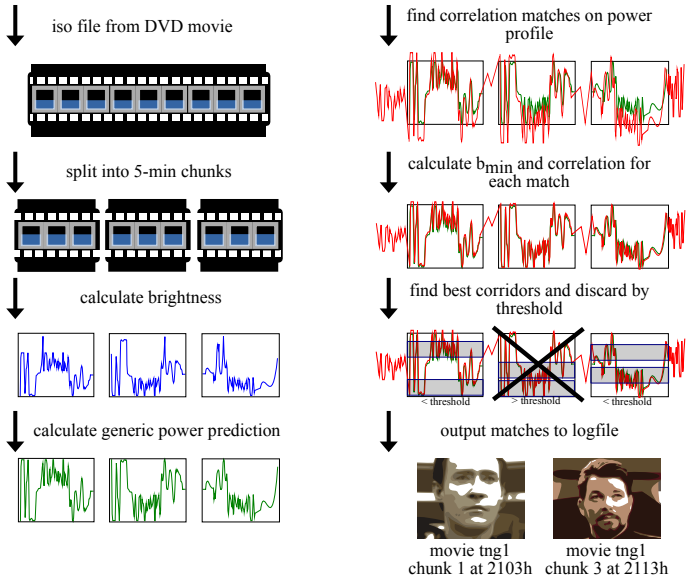

Fig. 8. Work-flow to detect chunks of a movie

G. Discoverable Video Material

During our experiments, some recorded television content such as German daily news[8] are not identifiable due to: 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 backlighted LCD TVs and the fluctuation of power consumption therefore can not be detected (detection is though still possible with CRT TVs: see Table I). Some other content such as the JAG[9] TV series has higher brightness level than typical b_{min} values: detection is harder when the shows is played on LCD TVs. Having gained some experiences with 653 content files and some days of recorded program broadcast, we could state that detection of movies produced for cinema projectors was almost always a feasible task while many TV studio productions (e. g. talk shows, news) are difficult or impossible to identify when played as recorded content. It is still rather easy to determine which TV broadcast consumers are watching since we only have to correlate the power curve to the content of a few dozen live stations and there is no need to search for a match along the time axis as the timing is provided by the program. The second or third rare dark scene (with brightness $b < b_{min}$) is then sufficient to identify the station on the power curve (idea: the first matching scene could be a coincidence: e. g. viewer interaction with TV teletext).

H. Other Television Models

Experiments described in the previous sections were performed on a home LCD TV[10] equipped with dynamic backlight enhancing technology. To support our claim that content

[8] *ARD Tagesschau*: daily German news broadcast at 8 p.m.
[9] Director: Donald P. Bellisario, air date: 1995 - 2005
[10] Panasonic model number TX-L37S10E

TABLE I
LIST OF TELEVISIONS

Manufactor	Model Nr. technology	$watt_{\min}$	$watt_{\text{diff}}$	b_{min}	correl. $M1$[12]	$M2$[13]	$M3$[14]	TV show[15]
Panasonic	TX-L37-S10E LCD	~ 45	~ 70.0	$26-58$	0.9599	0.9283	0.9487	< 0.85
LG	47LH4000 LCD	~ 65	~ 1.5	$25-84$	0.9458	< 0.85	< 0.85	< 0.85
Orion	TV32FX-100D LCD	~ 100	~ 3.0	$50-232$	0.8958	0.9402	0.9326	0.8989
Panasonic	TX-P50S-20E Plasma	~ 45	~ 160.0	$81-92$	0.8722	0.9510	0.8871	0.8933
Sony	KDL46EX-505 LCD	~ 170	–	–	< 0.85	< 0.85	< 0.85	< 0.85
Telefunken	Cinevision 32 CR tube	~ 60	~ 50.0	$58-153$	0.8833	0.9454	< 0.85	0.9283

identification is in general possible, we have performed experiments on other TV models as well.[11] We connected a out-of-box smart meter (meter Q3D from Easymeter GmbH) with specific TV sets.

For those tests not conducted with the operational smart meter, all data transmission took place directly between the smart meter and a connected notebook computer. We have played a total of 7 films and 2 TV shows. Table I contains the detailed test results relating to 3 movies and 1 TV show.

The successful test results affirm our belief that movie/TV content identification via fine-grained smart meter data is possible (with the exception of the Sony LCD test case). We also would like to point out that content identification using a cathode ray tube or a plasma display is also possible if very bright frames are taken into consideration. For some LCD TV sets not supporting dynamic backlighting (see sets *LG*, *Orion* in Table I) the power consumption difference between light and dark frames is rather small so content identification might become infeasible, especially if other appliances generate some noise to the smart meter data.

I. Determining the Optimal Values

To optimize the set of parameter values (minimum correlation that qualifies a matching content, corridor height, threshold for discarding a match), a series of test runs for varying parameters are performed (Table II). Starting from values suggested by the first experimental results, different values were subsequently tried. The number of false positive and correct identification (last two columns) are recorded. In conclusion, the combination $0.85, 50\%, 5\%$ (second last row) seems to produce the best identification rate. This means that the 5 minute chunks are compared to the power curve using a sliding window along the power axis and a preliminary match is declared when the correlation exceeds 0.85. The match is discarded if more than 50% (threshold) of the values fall in two corridors each having a height of 5% of the whole interval. Two optimal corridors maximizing the corridor coverage are to be identified for each match. The discarding is done twice: On the predicted power consumption values and on the measured power consumption values.

The goal of the elimination process is to prevent false positive matches but it also leads to discarding of about half of the correct hits (Table II shows the results for 12 movie chunks not being discarded by predicted power values). Since a movie consists of 18 or more 5 minute chunks, this discarding procedure is applicable in a real-life scenario of content identification. The parameter combination $0.85, 50\%, 5\%$ provides the identification of 11 out of the 12 chunks while one false positive match was logged. We used a collection of 653 content files to search for content matches.

[11]The authors wish to thank graduate student Stephan Brinkhaus BSc. who conducted various tests with the smart meter on several TV sets and other appliances.

TABLE II
CONTENT IDENTIFICATION OF 12 MOVIE CHUNKS WITH DIFFERENT PARAMETERS

Correlation	Thresh.	Height	PC Codr	PC Height	False Positive	Identification
0.9	0.8	0.10	0.8	0.10	5	9
0.85	0.8	0.10	0.8	0.10	12	11
0.8	0.8	0.10	0.8	0.10	69	12
0.9	0.7	0.10	0.7	0.10	2	9
0.85	0.7	0.10	0.7	0.10	4	11
0.8	0.7	0.10	0.7	0.10	34	12
0.9	0.6	0.10	0.6	0.10	0	7
0.85	0.6	0.10	0.6	0.10	0	8
0.8	0.6	0.10	0.6	0.10	1	8
0.9	0.6	0.05	0.6	0.10	0	8
0.85	0.6	0.05	0.6	0.10	1	9
0.8	0.6	0.05	0.6	0.10	5	9
0.9	0.6	0.05	0.6	0.05	2	9
0.85	0.6	0.05	0.6	0.05	6	11
0.8	0.6	0.05	0.6	0.05	39	11
0.9	0.5	0.05	0.5	0.05	0	7
0.85	0.5	0.05	0.5	0.05	0	8
0.8	0.5	0.05	0.5	0.05	3	8

J. False Positives with Other Appliances

In order to get some consolidated findings regarding false content identification, we used our scripted content identification work-flow (depicted in Figure. 8) to search for content in several 24h-periods, in which power metering data are concurrently generated by different household appliances. Four persons were living in the household and using the appliances. We used our available set of 653 content files – split into 5-minute chunks – to search for film material. We count a (false positive) hit for every match having a correlation of at least 0.85, and there are 35.5 hits per 24 hours (see Figure 9 for example hits' log entries).

```
INFO  First correlation discard threshold: 0.0
INFO  Second correlation discard threshold: 0.85
INFO  Corridor discard threshold: 0.5
INFO  Corridor height: 5
INFO  Power consumption corridor discard threshold: 0.5
INFO  Power consumption corridor height: 5
INFO  Analyze log file "discovergy-Raw-2011.11.19_0100-2011.11.19_2359.csv"
INFO  *(_csv_5Min.Saw.6.1080p.mkv.csv) at 07:14:50
        cor =          0.8582734470331655      bMin =          40.0
        delta =        2.160000520199958       distance =      0.40085441550795814
        corridor =     0.4533333333333333      pcCorridor =    0.31    [...]
INFO  *(_csv_5Min.Spaceballs.mkv.csv) at 11:06:43
        cor =          0.8554506191367586      bMin =          28.0
        delta =        316.08000034434         distance =      57.05935049506766
        corridor =     0.42                    pcCorridor =    0.3333333333333333
INFO  *(_csv_5Min-Jackie.Chan.-.Action.Hunter.avi.csv) at 12:54:07
        cor =          0.8794719071839578      bMin =          48.0
        delta =        4239.1799992422         distance =      466.5877682097706
        corridor =     0.25333333333333335     pcCorridor =    0.36    [...]
```

Fig. 9. Log file clipping showing false positive matches on other appliances

Analyzing the hits shows that these can easily be identified as false positive matches because the power curve does obviously not reflect television operation. See Figure 10 as an example showing such a false positive match: the power consumption difference of more than 4000 watts is too high

for being generated by a TV set and the curve shape does not obey the shape of prediction.

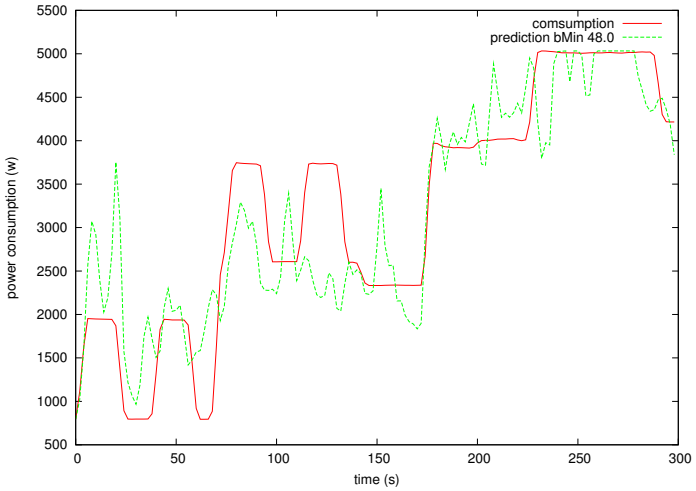

Fig. 10. Example of a false positive match

To avoid time-consuming manual discarding of the hits, a straightforward strategy to identify content would be to count only findings consisting of more than one match of a 5-minute chunk. A log entry showing two corresponding chunks of the same content (like the example of Figure 8: chunk 1 at 21:03h and chunk 3 at 21:13h) ruled out false positives on other appliances during the experiments. Note that we only used 653 content files for our experiments; a forensic investigator who is searching for copyright-protected content of all movies and TV productions ever been produced might have to solve a more challenging false-positive-problem. We did not have sufficient content files to reach a proper assessment on the feasibility of a forensic software.

VI. Conclusion

Smart meters are able to become devices that monitor the behavior of their customers. The personal privacy invasion is obvious if the smart meter data are available to malicious parties or being used by members of the same household to spy on each other.

A new generation of smart meters generating high-resolution energy consumption data could henceforth cause new potentials concerns regarding consumers' privacy sphere. We have demonstrated that particular information available on appliances in the household via its detailed power profile allow a fine-grained analysis of the appliance's behavior. Taking measurements at an interval of two seconds is sufficient to enable the identification of a television program or audio-visual content if favorable conditions are in place (e.g., no major interference of other appliances for minutes long). Our research has shown that the electricity usage profile with a $0.5s^{-1}$ sample rate leads to a *invasion* into a person's private sphere regarding his TV watching habits. Five minutes of consecutive playing of a movie is in many cases sufficient to identify the viewed content by analyzing the smart meter power consumption data. While we did not have sufficient content files to generate affirmative statistical evidences that can lead to a forensic software that will police every copyright infringement material. Our paper shows that there is at least a major privacy issue regarding content identification via a smart meter.

Our investigation also reveals that a smart meter transmits data via the Internet unsigned and unencrypted. This is a major throwback in data integrity and consumer privacy. This technical flaw could be attributed to the startup nature of the installer company who is in a phase of service development and product quality definition. It nevertheless proves that a minimum regulatory requirement regarding smart meter data protection standards need to be defined and fulfilled, before a meter becomes fully operational and capable of preserving a user privacy.

References

[1] The basics of tv power. http://reviews.cnet.com/green-tech/tv-power-efficiency/, April 2010.

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

[3] A. Rial and G. Danezis and M. Kohlweiss. Differential private billing with rebates. Technical Report MSR-TR-2011-10, Microsoft Research, Feburary 2011.

[4] Gergely Ács and Claude Castelluccia. Dream: Differentially private smart metering. *CoRR*, abs/1201.2531, 2012.

[5] G.Danezis A.Rial. *Privacy-Preserving Smart Metering*, MSR-TR-2010-150.

[6] Miro Enev, Sidhant Gupta, Tadayoshi Kohno, and Shwetak N. Patel. Televisions, video privacy, and powerline electromagnetic interference. In *ACM Conference on Computer and Communications Security*, pages 537–550, 2011.

[7] Flavio D. Garcia and Bart Jacobs. Privacy-friendly energy-metering via homomorphic encryption. In J. Cuellar et al., editor, *6th Workshop on Security and Trust Management (STM 2010)*, volume 6710 of *Lecture Notes in Computer Science*, pages 226–238. Springer Verlag, 2010.

[8] G.W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.

[9] Marek Jawurek, Martin Johns, and Florian Kerschbaum. Plug-in privacy for smart metering billing. In *PETS*, pages 192–210, 2011.

[10] K. Kursawe. Some Ideas on Privacy Preserving Meter Aggregation. Technical Report ICIS–R11002, Radboud University Nijmegen, January 2011.

[11] Klaus Kursawe, George Danezis, and Markulf Kohlweiss. Privacy-friendly aggregation for the smart-grid. In *PETS*, pages 175–191, 2011.

[12] 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.

[13] Andres Molina-Markham, Prashant Shenoy, Kevin Fu, Emmanuel Cecchet, and David 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.

[14] Klaus Mueller. Gewinnung von Verhaltensprofilen am intelligenten Stromzähler. *Datenschutz und Datensicherheit - DuD*, 34:359–364, 2010. 10.1007/s11623-010-0107-2.

[15] Ronald Petrlic. A privacy-preserving concept for smart grids. In *Sicherheit in vernetzten Systemen: 18. DFN Workshop*, pages B1–B14. Books on Demand GmbH, 2010.

[16] 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.

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

[18] Stephan Renner. *Smart Metering und Datenschutz in Oesterreich*, DuD 2011.

[19] Robert Scott West, Huub Konijn, Simon Kuppens, Nicola Pfeffer, Quint van Voorst Vader, Yourii Martynov, Tewe Heemstra, and Jan Sanders. Led backlight for large area lcd tvs. In *10th International Display Workshops (IDW)*, 2003.