

Measuring Voter Lines

Carsten Schürmann
Jian Wang

**Copyright © 2015, Carsten Schürmann
Jian Wang**

**IT University of Copenhagen
All rights reserved.**

**Reproduction of all or part of this work
is permitted for educational or research use
on condition that this copyright notice is
included in any copy.**

ISSN 1600–6100

ISBN 978-87-7949-356-8

Copies may be obtained by contacting:

**IT University of Copenhagen
Rued Langgaards Vej 7
DK-2300 Copenhagen S
Denmark**

Telephone: +45 72 18 50 00

Telefax: +45 72 18 50 01

Web www.itu.dk

Measuring Voter Lines

Carsten Schürmann, IT University of Copenhagen
Jian Wang, IT University of Copenhagen

We present an automated data collection technique called *white boxes*- or simply *wb*-technique, that is designed

- to analyze the voter behavior in polling stations, which includes the measurement of arrival and waiting times, the determination of arrival frequency and multiple arrival;
- to assist the management of polling places to make decision regarding the distribution of resources, and to identify areas for future improvement;
- to collect historical data to evaluate the impact of redistricting and gerrymandering on the quality of the voting experience;
- and to provide hard data to guide the political decision making process with respect to the choice of voting technologies.

1. INTRODUCTION

The question of how to improve the voters' experience when going to the polls is of central importance to many electoral management bodies (EMBs). Good voter experience is often associated with an efficient and professionally organized electoral process, which in turn is expected to lead to elevated levels of voter participation, voter satisfaction, and also trust in the overall election and its outcome. In a recently published report by the US Presidential Commission on Election Administration on the American Voting Experience [Bauer and Ginsberg 2014], for example, one of the recommendations calls for *state-of-the-art techniques to assure efficient management of polling places, including tools the Commission is publicizing and recommending for the efficient allocation of polling place resources*. Such state-of-the-art techniques can be roughly divided into two categories: techniques that render polling places more efficient and techniques that measure polling place efficiency. Denmark, for example, uses digital voter registration systems to increase efficiency: By presenting a voter registration with a machine readable barcode, voters can be quickly and efficiently checked of the electoral roll.

This paper presents a state-of-the-art technique for analyzing the efficiency of a polling place by measuring the times voters are present in and around a polling place. This analysis technique is called the *white boxes*- or simply *wb*-technique. The results of the *wb*-technique are datasets from which we can infer information, for example, when a voter arrived at the polling place, when he or she left. Using statistics, we can deduce valuable information that can assist administrators to render polling places more efficient. Examples of the effects of such a method include: Reallocating resources from one polling station to another in the case that there is a demographic change; purchasing new equipment, such as new ballot boxes, additional registration desks, voting machines or curtains, with the goal to shorten long waiting times; it may also lead to restructuring the layout of polling places to improve flow; or collecting statistical information to quantify voting culture, for example, how many voters skip out of a voting line and leave, how many return again, and how many remain in the polling place after they cast the ballot to wait for friends or chat with acquaintances.

With every redistribution of resources, there is some risk that the quality of the election decreases instead of increases. There are reports from US elections, for example, that after gerrymandering some polling places experienced extremely long waiting times, such as for example, during the 2012 presidential election in Richland County [Buell 2013]. The data that we collect at polling places can be used to understand these problems better and evaluate the effectiveness of counter measures. It can be used to justify expenses towards more efficient polling place administration and to disarm arguments based on circumstantial evidence, for example unsubstantiated complaints about excessive waiting times.

In the US, there is a critical awareness that such data is invaluable. Per recommendation of the CalTech/MIT voting project, several precincts have already collected or are planning to collect queuing data the old fashioned way, i.e. by having election officials count the number of people standing in line in regular intervals; or by handing out pieces of paper to voters that are stamped with arrival and eventually also the departure times; and by tracking individuals throughout the voting process, marking the arrival times at individual service points. In this paper we show that our *wb*-technique, is on all accounts superior to the manual counting: It is more reliable in that it is virtually free of human error; it is more consistent as information is continuously recorded; it permits (at least in theory) the reconstruction of the path of individual voters through the polling station by trilateralization and it is unobtrusive, because sensors can be installed in the polling place out of sight of the polling officials and voters. Furthermore, in this paper we show, that the *wb*-technique and the manually collected data can be correlated. In this paper we show, that *it is possible to replace the manual collection of queuing data by a technological solution, while improving the accuracy of the measurement.*

This paper is organized as follows. In Section 2 we describe the white box-technique. The hardware and software that we chose to implement the technique are described in Section 3. Security and privacy considerations are discussed in Sections 4 and 5, respectively. We describe a pilot study with this technology that we conducted during the 2015 Danish parliamentary election in Section 6. We deployed the technology in five polling stations including three in Copenhagen and two in Aarhus. For one polling place in Copenhagen, Holbergskolen, we conducted CalTech/MIT style manual collection of inflow data, which we use to evaluate the quality of our method. Finally, we assess results and describe our preliminary findings in Section 7. We conclude that the white box method provides a precise and accurate information to measure waiting times in polling places.

Legal Concerns. According to Danish data protection agency *Datatilsynet*, the media access control (MAC) address is considered sensitive (*personfølsom*) information, which means it is protected by data protection laws. Recording of this information is only legal if permitted by *Datatilsynet*. The DemTech project has received permission from the Danish data protection agency to record this data for scientific purposes.

2. WHITE BOXES-TECHNIQUE

The basic idea behind this technique, is that mobile phones send out wireless packets that can be recorded by a sensor for future analysis. Abstractly, we can describe the data collected this way by as a set of observations O , that records wireless packets in the form of pairs

$$(p, t) \in O$$

where p denotes the identifier of a mobile phone and t the time at which a wireless packet was observed. The sensors that record the packets sent out by mobile phones have only limited range. Therefore, it will become necessary to deploy multiple sensors for one polling station, which entails, that we will also have to combine multiple sets of observations O_1, \dots, O_n for our analysis. To compute for how long a mobile phone, aka voter was present at a polling place, we have to combine the sets of observations and compute the following presence relation:

$$(p, s, e) \in \text{Presence}(O)$$

where s (e) refers to the earliest (latest) time when (p, s) ((p, e)) was recorded in O . In other words, we can compute precisely for how long each device stayed in a polling place — by subtracting s from e .

Noise: In our pilot study that we describe below in Section 6, we have observed that empirically collected data sets contain noise. This noise may be due to other devices emitting wireless packets, such as, for example, routers that are installed in the building hosting the polling station, mobile phones of voting officials or people passing by without voting, or mobile phones of voters that run out of batteries. We remove the noise from the dataset using common statistical methods, in

Technology	Range
RFID tracking	< 1m
Bluetooth tracking	ca. 10m
WiFi tracking	< 100m
UTMS, GSM	800m – 40km

Fig. 1. Tracking Technologies

particular, one standard deviation, which is defined as follows.

$$\sigma = \sqrt{E[(T - \mu)^2]}$$

where $\mu = E[T]$ computes the expected value. Data outside the range of $[\mu - \sigma, \mu + \sigma]$ is treated as noise.

Accuracy: Each sensor, will accurately record any wireless packet observed within its range, assuming that the local clock of the sensor is configured correctly. This has two consequences. First, if the waiting queue extends outside the range of a sensors installed, it will not “see” the end of a line, and conversely, a mobile phone may be detected while the voter is still in progress of enqueueing. In the former case, it is important that the voting officials can predict where the line will form and then use sufficiently many sensors to monitor the queue. In the latter case, it might be possible to compensate for this mathematically, but for the purpose of this paper we have decided not to look into it any further. This means, that in comparison with manually recorded waiting data, our graphs are slightly more conservative.

Completeness: We note, that not every voter will carry a mobile phone that emits packets. In our experience, this is not problematic, because of the way lines are formed, as long as sufficiently many people in line carry such a phone. Our data shows, that most commonly every third to fourth voter (in extreme cases every eighth voter) carries such a phone. Naturally, the more mobile phones are present in a voting place, the more accurate our estimates for waiting times will become.

We also note, that the frequency with which mobile phones emit packets depends very much on the phone’s operating system. This is because the probe frequency, for example, for WiFi ranges from 4 packets per minute to less than one.

We anticipate a future complication with the *wb*-technique when phones come pre-configured with an anonymization enabled. This technique hides the device identifier of the phones during the probing phase. However, this is currently not a concern, as only few phones have this feature enabled [Motorola Solutions 2014], in part, because it is so inconvenient to enable it. All notification services, email, messages, etc. must be disabled for this feature to work.

3. IMPLEMENTATION

Next, we describe our design choices for implementing the *wb*-technique, in preparation for a real election.

3.1. Hardware

The first design choice is about which kind of wireless packets we intend to record. There are several options as outlined in Figure 1. Among these options, we have selected WiFi tracking: Most smartphones and smartwatches have active WiFi on their phones, simply because of convenience. This way the phone will try to connect automatically to an access point at home or in the office. When trying to connect, the smartphone periodically emits a probing packet, which we will then record. The range of WiFi phones in the open is around 100m, indoors between 10 and 20m and is therefore better than Bluetooth, which has a much smaller range. In addition many more people use their smartphones to connect to access points than have Bluetooth enabled. We briefly also considered RFID tracking, but discarded this idea because of low range and security concerns. Recording UMTS or GSM signals is in general illegal.

WiFi enabled devices usually transmit on 11 different channels. When trying to connect to an access point they transmit a broadcast packet on all channels, which means that for our system, it is sufficient to listen only to one channel (channel 1) which uses a frequency of 2412 MHz. Note, that the mobile phone has also Bluetooth enabled, there might be some interference between WiFi and Bluetooth.

The second design choice is about the particular technology that we shall use to record wireless packets. There are many possibilities, ranging from small computers (for example, Raspberry Pi) to routers. In preliminary experiments, we used a TP link travel router TP-MR3020.¹ These routers have the advantage that we can install OpenWRT (a Linux derivative) on them. They also come with a USB port, which we use to connect a USB drive to record the data persistently, which we found worked well. We also used USB drives in an earlier pilot Danish Municipal election 2013, without any problems.

An alternative technology are 3G antennas, that we used for the European parliament election 2014, where we encrypted and transmitted the data directly to DemTech's server. This had the advantage that in principle, we could provide online queuing data information, however, the 3G connection to the server was less reliable than expected, so that our data collection was spotty at best.

One drawback of these sensors was that they needed to be connected to an external power source, which means that the location for the sensors in a polling place was largely determined by the location of power outlets.

For the 2015 Danish Parliamentary election we therefore started to look for another technology and we chose the TP link travel router TP-MR13U² and 16GB SanDisk Cruzer Fit USB Flash drives. This router comes with an embedded battery large enough to power the sensor for 48 hours straight, ample of time to be deployed during an election. Just like with the TP-MR3020, OpenWRT can be installed on this sensor and they can be programmed to start and stop collecting data at precise time points. Practically speaking a polling station can be prepared the day before the election, and the sensors can be collected after the polling station closes.

3.2. Software

To prepare the routers, we flushed the router's firmware and replaced it by OpenWRT³, a version of the Linux kernel that is optimized for network traffic management. Next, we set the wireless interface of each sensor into passive mode, which means that the sensors can only listen. One can think of the sensors as microphones, recording all wireless traffic around them.

OpenWRT supports many of the common networking tools, in particular, tcpdump⁴ a tool that we configured to record all wireless traffic. We configured the sensors in such a way that they would automatically start recording packets at 08:00 and stop at 21:00 on the election day.

4. SECURITY

We review the security of our system from the point of assets and vulnerabilities [Basin et al. 2009] in the polling places. In our experiments, the physical assets are the routers and their respective USB drives. The physical assets' integrity can be violated due to unexpected damage to the sensors, for example, through water, fire, or theft. As logical assets we refer to the router's software and the data collected on the USB flash drive.

Regarding the protection of the physical assets, there is little we can do beyond physical securing them, because the sensors are not monitored. During the experiments described below, we placed the routers in difficult to access places and out of the voters sight.

¹http://www.tp-link.com/en/products/details/cat-4691_TL-MR3020.html

²<http://wiki.openwrt.org/toh/tp-link/tl-mr13u>

³<https://openwrt.org/>

⁴<http://www.tcpdump.org/>

We therefore direct the focus of this security analysis on the integrity and security of the logical assets. Here, we even assume that the sensor itself is under the adversary's control, and therefore as the primary security mechanism we protect all data on the USB flash drive, by encrypting the relevant partition on the drive. The key for decrypting the flash drive was not stored on the sensor, but kept in our office.

As a secondary security mechanism, we restrict the attack surface of each sensor: the only way to access it is by attaching network cable and logging in through ssh.

The only personnel authorized to access are the system administrators, who are usually not in field. A system administrator has all keys and thus has the access to all critical data. We argue that the system is secure keeping in mind that an attacker could easily record the same wireless traffic without having to breach the security measures that we have put into place.

Next, we'll address the protection of the collected data. Each recorded packet contains identifying information about the identity of the mobile phone, the so called MAC address. Even though it is possible to reset the MAC address of any such device, it is uncommon for users to do so. Note, that for the purpose of this work, it is important, that

- MAC addresses are properly anonymized to protect the identity of the voter,
- only data relevant to our experiments is kept, including anonymized MAC addresses, time stamps, and signal noise ratios, and
- the anonymization of mac addresses is deterministic, because we need to correlate packets across multiple sensors.

For practical purposes, during the voting day, all data is recorded. Any anonymization and analysis of the data is done post-election.

The data collected during this pilot must be classified as "sensitive" because MAC addresses are considered personal data. To do an appropriate security analysis of our design, we consider which possible adversary may be interested in stealing the data we collect. An adversary may want

- to steal data out of curiosity,
- to gain access to recorded information,
- to discredit the election commission,
- or to interrupt the pilot.

It is therefore important to take the necessary technical and operational security precautions. However, in general, the security threat is at most moderate, as many polling places are within the range of wireless routers that could be reprogrammed to collect similar kinds of information. Every smart phone can be programmed to do the same.

We use disk encryption (AES based, SHA1, 256 bit) to deter attackers and to secure the data stored on the USB drive on the sensor. The disk encryption subsystem is part of the Linux kernel (`dm -crypt`). This way, we can guarantee that all data on the USB drive is encrypted and thus rendered useless for an attacker stealing it.

5. PRIVACY

The MAC address for a network interface, is factory preset for each networked device. This includes smartphones, but also standard networking hardware that can be found in laptops and other computers. A MAC address contains 6 octets, where the first 3 octets identify the organization that issued this MAC. Together with the rest 3 octets, it is possible to identify the device that owns the MAC, which may lead to a breach in privacy. For most devices, in particular Apple's iPhone or Android phones, the MAC address can be reset by the user to any randomly chosen 6 octets⁵. For example `c2:31:9d:d0:30:e8` is a valid MAC address. As each octet is 8 bit, there are 2^{48} unique MAC addresses. Standard anonymization techniques, for example, by computing the SHA256 digest for MAC from a voters phone provide only low levels of security, because a simple dictionary attack

⁵Some special MACs are reserved

would allow an adversary to relate the anonymized identifier to the original MAC address. So we chose to use HMAC (keyed-hash message authentication code) to hide the MAC before we publish the raw data, where the HMAC key protects against the dictionary attack.

For reasons of reliability, when we use tcpdump to capture the wireless packets, we do not filter out the packet header during recording. Instead, we stored the whole packet, and remove the packets body offline. The secrecy of each packet's content depends on the type of security method used for wireless communication, for example WEP, WPA, or WPA2. In addition, most voters will not be able to connect to any of the local networks in the polling station, which means all the captured packets will only be handshake beacons, instead of payload carrying network packets, that arise when checking email or surfing the web.

Our security and privacy measurements are not designed to protect against inside attacks, which means if the keys are misused by an inside attacker, the attacker may gain free access to the data on the USB drive and recover the original MAC addresses from the anonymized ones. If the attacker is in possession of the secret key used to compute the HMAC, he will be able to brute force the original MAC address.

6. FIELD STUDY

In this section we describe the 2015 pilot project, where we recorded network traffic using the white boxes method and recorded them manually following the CalTech/MIT style.

The general parliament election was held in Denmark on 18 June 2015 from 09:00 to 20:00 to elect the 179 members of the Folketing (Danish parliament). We installed our white box technology in 5 polling places in two cities, Aarhus and Copenhagen, and deployed in total 18 white boxes with a minimum of 3 white boxes installed in each each polling station. Out of the 5 polling places, 3 were located in Copenhagen, which are Holbergskolen, Bellahøj Skole and Islands Brygge Skolen, and 2 were located in Aarhus, which are Møllevangskolen and Frederiksbjerg Hallerne.

6.1. Installation

The position for each white box in a polling place was carefully chosen, in an effort to ensure that all white boxes together cover the entire polling station and the area where there queues were expected to form. In some polling places we placed white boxes in such a way to achieve a higher level of redundancy, useful for the case when a sensor fails. As shown in Figure 2a, 3 sensors were placed in Holbergskolen polling place in Copenhagen, where the bottom two sensors are placed to record the mobile devices before voters go through the entrance (IND) and after they leave (UD). The small circles mark the registration desks, the first service point. After registration, the voters received a blank ballot and proceeded to the second service point, waiting for an empty voting booth, which are marked in Figure 2a by a crossed out circle. The queues formed alongside the building, along the dotted line in the picture.

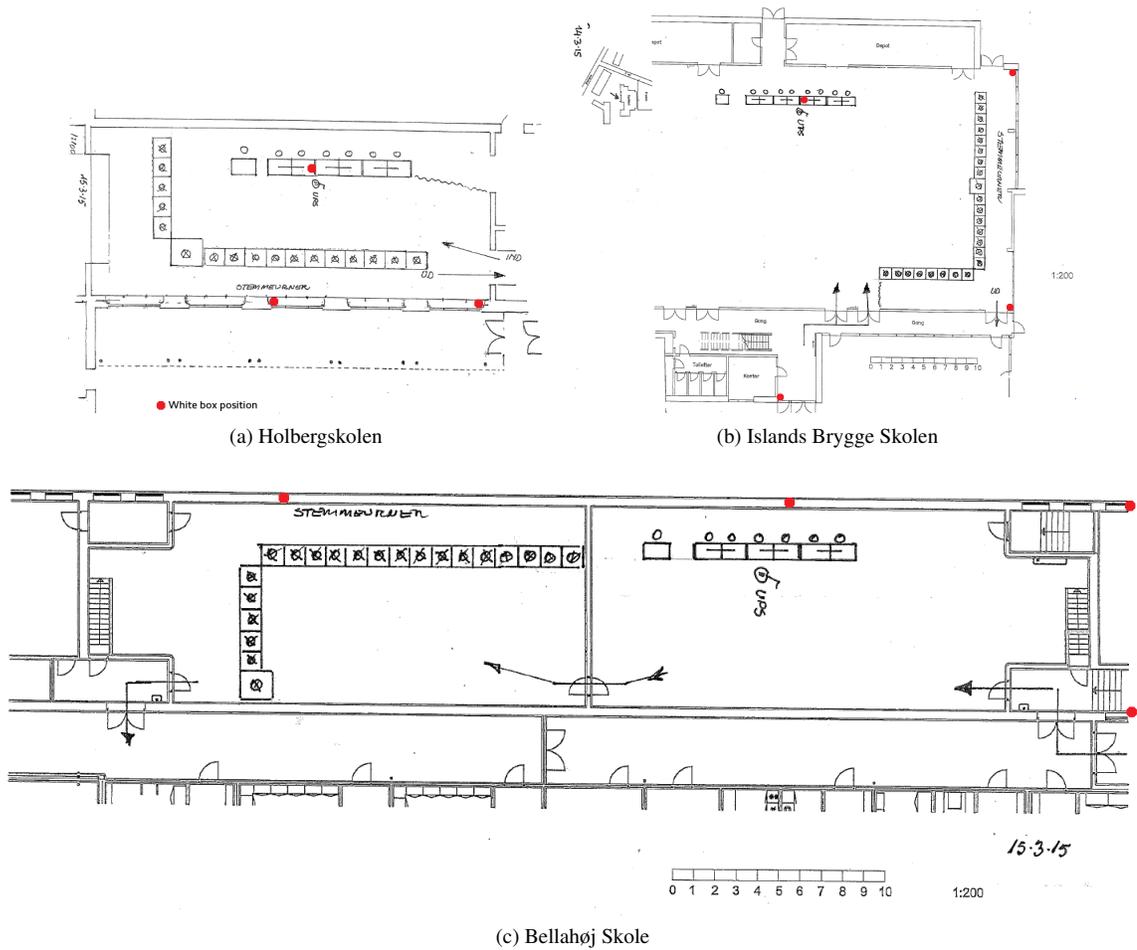


Fig. 2. Layout Maps of White Boxes in Copenhagen

On the day before the election, we deployed five teams who visited one of the five polling places each, and installed 3-4 sensors. The sensors were programmed to record from 08:00 to 21:00. The sensors were completely autonomous and did not require any servicing during election day. Each team collected the sensors after close of polling station and returned them to the lab for further processing.

6.2. Data Preprocessing

The data collected by each sensor consists of a list of packets. These lists contain several millions of packets each. Each packet is described by a time stamp, an anonymized MAC address, and a value representing signal strength.

Figure 3 depicts a plot of one of these sensors, before noise was removed. The x axis describes time and the y axis the identifier of the mobile phone, sorted in order of first appearance. As we can see, several sensors leave long lines to the right; this may be due to routers in proximity to the polling station being turned on, or other reasons.

First, we tweak the data a little:

- (1) We preprocess the data in the following format, summarizing all observations at times $t_1 \dots t_n$ of the same identifier into one tuple $(p, \{t_1, \dots, t_n\}, l)$. l refers to the duration an identifier was within scope of the polling station $l = t_n - t_1$.
- (2) We remove all tuples from this set, where $l = 0$.
- (3) We remove all tuples, where $60min < l$.
- (4) We remove all tuples, recorded $30min$ before the polling station opens, i.e. where $t_1 < 08:30$.
- (5) We remove all tuples, recorded $30min$ after the polling station closes, i.e. where $20:30 < t_n$.
- (6) We remove all tuples corresponding to *spurious* identifiers, where $l < 1min$.
- (7) We remove all tuples corresponding to *transient* identifiers, i.e. there exists an i , such that $10min < t_{i+1} - t_i$.
- (8) Finally, we apply the noise removal techniques described earlier, which significantly improved the overall quality of the data. With a running average (deviation) μ (σ) of the duration l for 10 minute intervals (empirically determined), we remove in addition all those tuples for which $l < \mu - \sigma$ or $\mu + \sigma < l$.

It is also conceivable to preprocess the data set even further: For example, for each wireless packet, we could determine the manufacturer information (before anonymization) and based on this information, we could decide to keep the tuple or remove it. This is possible, because the first 24 bits of a MAC address identifies the manufacturer as discussed in Section 5 (unless the MAC address was reset). There are online inverse MAC-address lookup services.

When we black-listed the five common router manufacturers, in particular CISCO, TP-LINK, DLINK, MicrosoftServer, and Netgear, and removed all packets from those devices from our data set, we observed, that our method produced nearly identical results. This means, that our data processing step effectively removes all black-listed devices automatically.

We also observed, that our sensors sometimes behave erratically. The sensor No.7's unprocessed data depicted in the Figure 3, for example, appeared to have malfunctioned as it stopped recording packets for about one hour. The damage could be mitigated, because the other sensors in polling places worked well. We remark, that this was also the only malfunctioning that we observed.

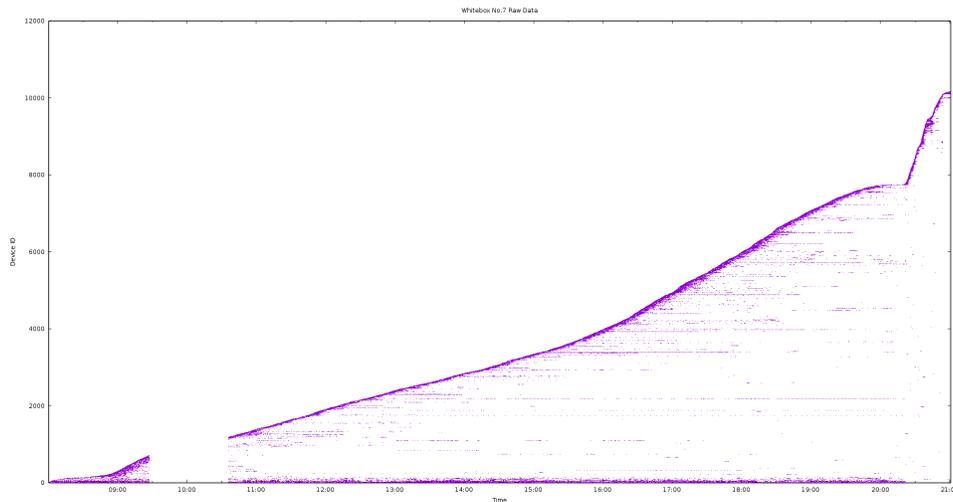


Fig. 3. White Box No.7 Data

6.3. Devices

In Denmark, it is common practice to publish statistical information about each election on the internet. The official election homepage [Danmarks Statistik 2015], for example, lists how many voters voted in each polling place. We use this information to approximate, how many voters carried a WiFi enabled mobile phone (see Figure 6.3). The overall penetration of smart phones in Denmark was 59% in 2013 ⁶.

Polling Places	Voters	Devices	Percentage
Holbergskolen	4997	1333	26.68%
Bellahøj Skole	5789	1942	33.55%
Islands Brygge Skolen	7931	3494	44.05%
Møllevangskolen	8043	3132	38.94%
Frederiksbjerg Hallerne	10578	3798	35.90%

Fig. 4. Device Capture Ratio

6.4. Queues

We show now how to interpret the data that we have collected using the polling station at Islands Brygge Skolen as example. The results for the other polling places is given in the Appendix A, Appendix B and Appendix C.

First, we visualize the data we have collected. Recall, that our data set consists of a set of tuples of the form (p, t_1, t_n, l) , where p stands for the anonymized identifier of each mobile phone, t_1 the point in time when the device was detected first, and t_n when it was seen last. Below we refer to this data set as \mathcal{D} . l refers to the total voter's waiting time (time spend in the polling place). The graph in Figure 20 visualizes this — the vertical axis ranges over device IDs ordered by t_1 , and the horizontal axis ranges over time. For each device, we mark t_1 by a blue and t_n by a red dot.

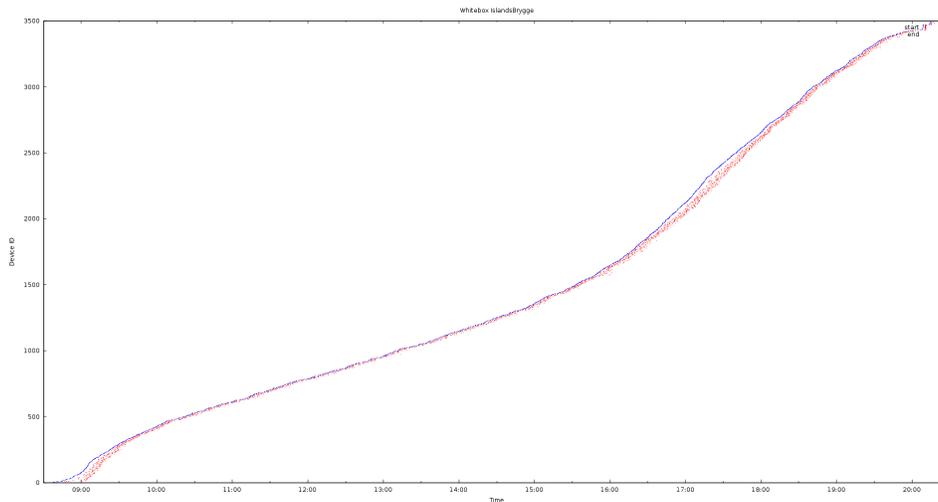


Fig. 5. Device Appearance in Islands Brygge Skolen Copenhagen

⁶https://en.wikipedia.org/wiki/List_of_countries_by_smartphone_penetration

Next, we describe two analysis techniques, for computing the average time that voters spent in a polling place. First, we use the method of averaging, which is described in Section 6.4.1, and second, we use a method based on Little's theorem [Little 1961] in Section 6.4.2.

6.4.1. Method based on averaging. With this method, we compute the running averages of the time spent in a polling place, using a sliding window of 10 minutes, 30 minutes, and 1 hour, respectively:

$$\mu_{len}(t) = \frac{\sum_{l \in L} l}{|L|}$$

where L is defined as $\{l | (p, t_1, t_n, l) \in \mathcal{D} \text{ and } t - \frac{len}{2} \leq t_1 \leq t + \frac{len}{2}\}$ and len is the size of the window. Figure 6 depicts a green, blue, and red graph for the polling place at Islands Brygge Skolen representing the running averages for a 10 minute, 30 minute, and one hour window, respectively. The horizontal axis denotes time and ranges from 08:30 to 20:30. The vertical axis denotes the average waiting time, in seconds.

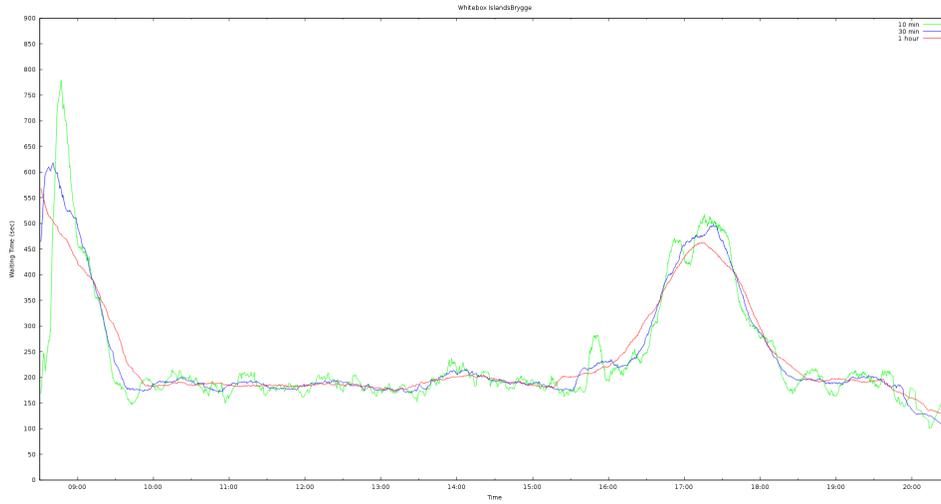


Fig. 6. Average Time (Averaging Method) in Islands Brygge Skolen Copenhagen

6.4.2. Method based on Little's theorem. In queuing theory, Little's theorem is widely applied. Little's theorem states [Little 1961]: "The average number of customers in a system (over some interval) is equal to their average arrival rate, multiplied by their average time in the system." Based on this theorem, we can compute the average time the voters spent in the polling station as

$$\mu_{len}(t) = \frac{V_{len}(t)}{\lambda_{len}(t)},$$

where len stands for the size of the window, as above. V_{len} is the average number of voters in the polling station at time t , and λ_{len} is the arrival rate also at time t .

The accumulated inflow values, outflow values, and their difference, which corresponds to the number of voters present in the polling station at a particular time, are plotted in in Figure 7. The grey, green, and blue graphs, correspond to inflow, outflow, and number of voters, respectively.

Figure 8 shows the average time voters spent in the Islands Brygge Skolen polling station for any particular point in time during the voting day. The color schemes are consistent with those used

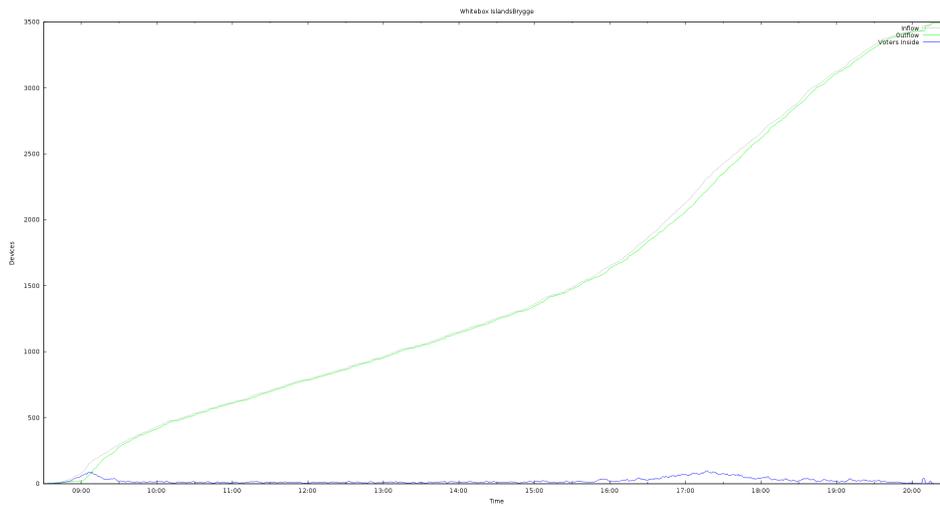


Fig. 7. Device Flow in Islands Brygge Skolen Copenhagen

above, green refers to a 10 minute window, blue to a 30 minute window, and red to a one hour window.

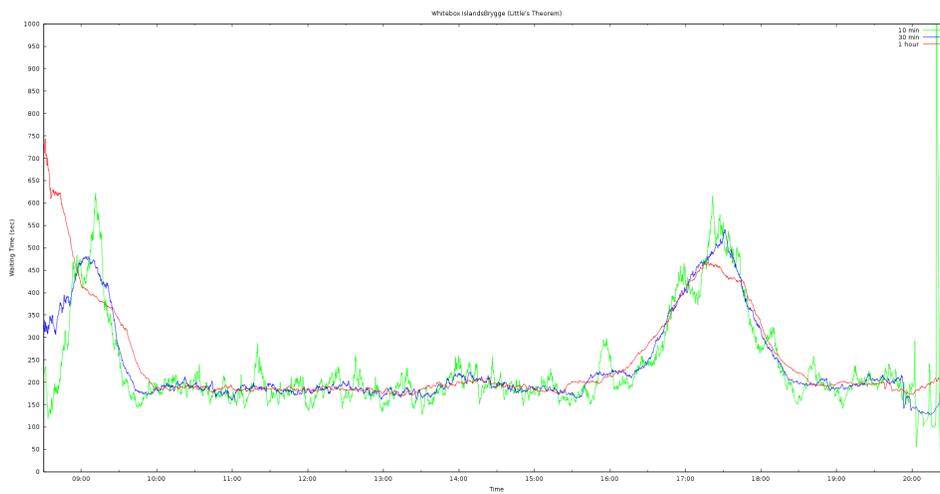


Fig. 8. Average Time (Little's Method) in Islands Brygge Skolen Copenhagen

6.5. Manual Count

In order to validate our method, we conducted in addition to the automatic method a manual count in parallel, following the CalTech/MIT recommendations, for assessing waiting times in polling places. In this section, we report which data we collected, and how to compare the white box method with the manual method. The CalTech/MIT recommendation proposes two different kinds of manual data collection.

- (1) The method asks an observer to follow and record the individual activities of each voter throughout the polling station. This includes, arrival times, for example, at the end of the queue, or different service points, and the departure of the voter from the polling station. Figure 9 depicts a fragment of the log that we manually recorded at the polling station, called Holbergskolen.

No.	Description	Check-in	Arrival registration	Leave registration	Enter booth	Leave booth	Ballot cast
16	green coat	10:51:40	10:53:00	10:53:15	10:53:24	10:54:09	10:54:14
17	purple jacket	10:59:45	11:00:35	11:00:51	11:00:57	11:02:00	11:02:11
18	lime jacket	11:10:06	11:10:58	11:11:20	11:11:30	11:12:22	11:12:28
19	red coat	11:19:10	11:19:24	11:19:40	11:19:48	11:20:26	11:20:28
20	white pants	11:28:10	11:28:30	11:28:44	11:28:46	11:29:16	11:29:20

Fig. 9. Example Data of Manual Count in Holbergskolen

- (2) The second method for data collection requires an observer to count the number of arriving voters in 10 minute intervals. The two leftmost columns of Tables 15–17 display this information for 4 different polling places. Due to the lack of observers, we were not able to record this data for the polling station at Frederiksbjerg Hallerne.

Next, we discuss the comparison between the white box method and the manual methods.

6.5.1. Queues. Here, we compare the data that we have collected in Holbergskolen using the white box method to Figure 9 above. From this data, we compute the time that each voter spent in the polling place, by subtracting “check-in time” from the “ballot cast time”. Using the averaging methods, we compute the running averages using a window size of 30 minutes and one hour, respectively. We remark, that window size of 10 minutes did not provide usable information, in part because of lack of observations. These two graphs are depicted in Figure 10 and Figure 11, using the colors light blue (30 minutes), and gray (1 hour). The remaining three plots (green, blue, and red) display the white box data for this polling place using the averaging method (see Section 6.4.1) and Little’s method (see Section 6.4.2).

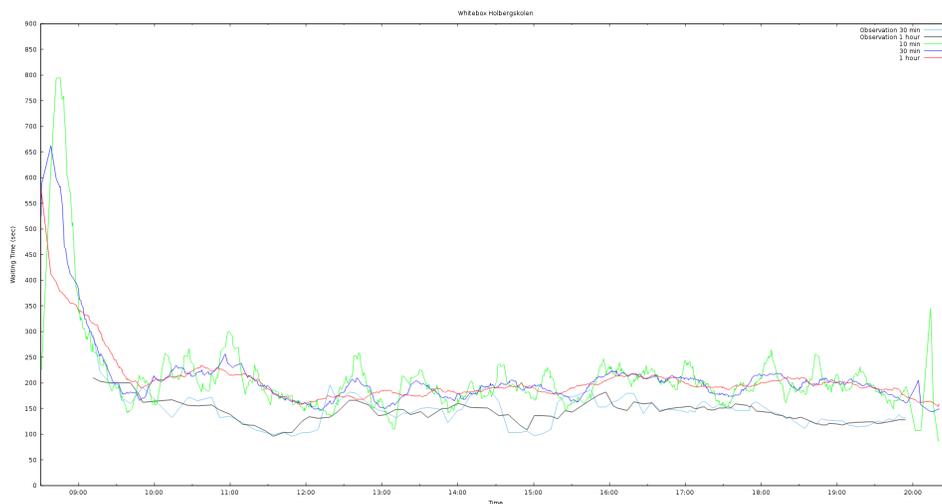


Fig. 10. Comparison of White Box (Averaging Method) and Manual Count in Holbergskolen Copenhagen

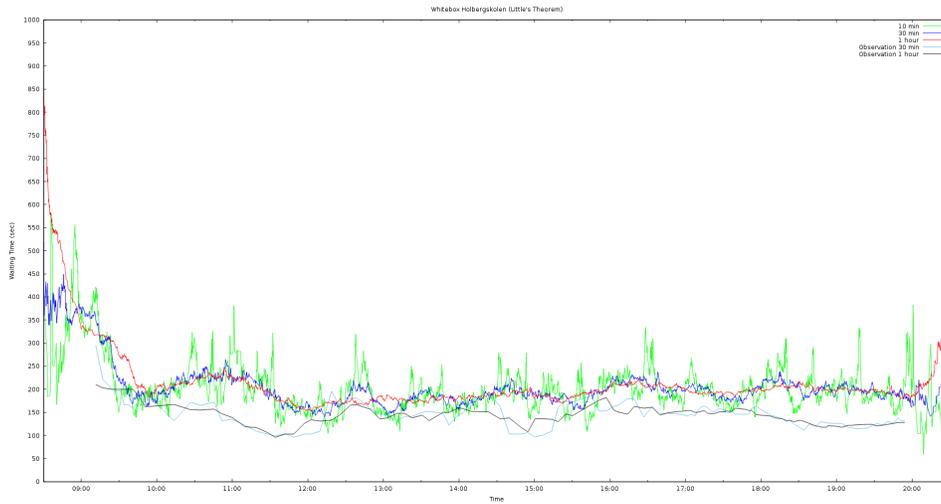


Fig. 11. Comparison of White Box (Little's Method) and Manual Count in Holbergskolen Copenhagen

6.5.2. *Devices.* Here, we compare the second type of data that we have collected manually with the data recorded using the white boxes. There is only one polling station, namely Holbergskolen, for which we recorded inflow manually during the whole day. Figure 12 depicts the inflow and outflow, recorded by our sensors in green and gray, respectively, and the inflow recorded manually in red. Note, that we did not record manually the outflow of voters from a polling station.

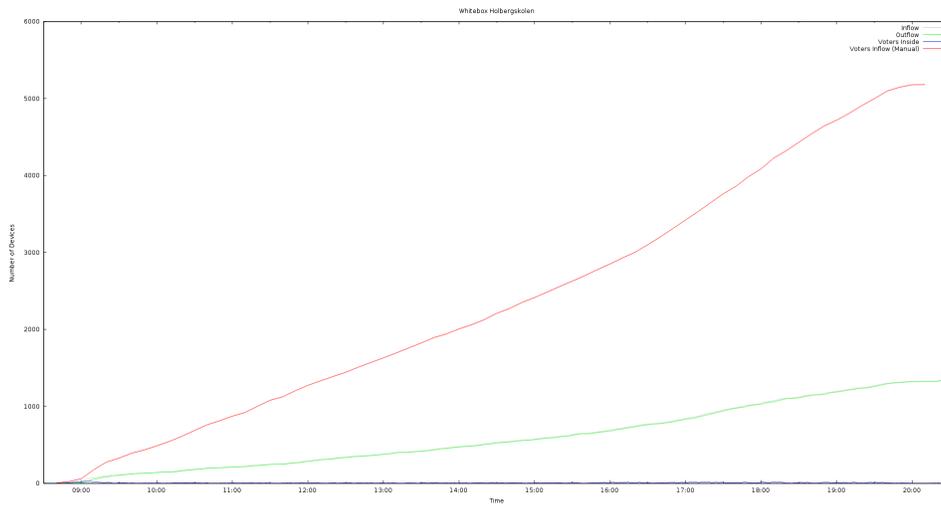


Fig. 12. Device Flow in Polling Station Holbergskolen Copenhagen including manual count

Figure 13 presents the same information but in a slightly different form. The red plot describes the average percentages of voters who carried a mobile phone, in 10 minute intervals.

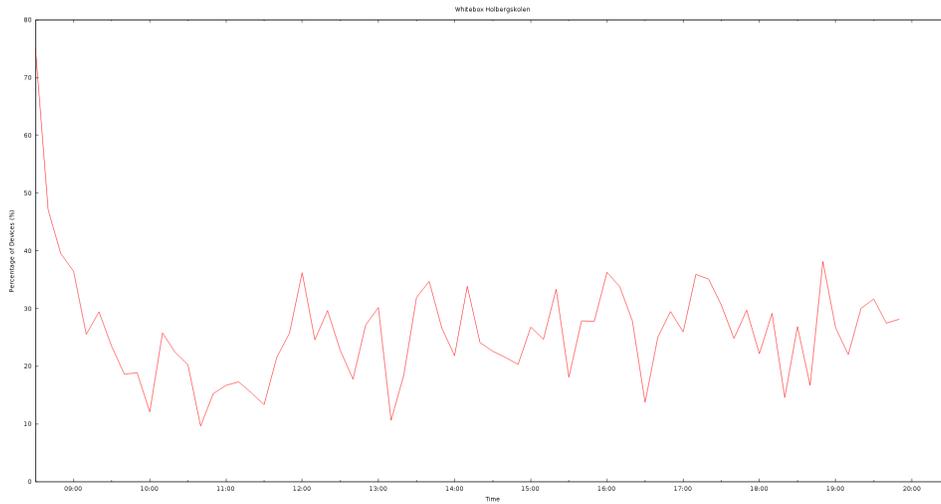


Fig. 13. Device Capture Ratio from 8:30 to 20:10 in Holbergskolen Copenhagen

In addition to the inflow collected from Holbergskolen (Figure 14), we also recorded the inflow for one hour period between 16:00 and 17:00 for three further polling places, including Bellahøj Skole (Figure 15), Islands Brygge Skolen (Figure 16), and Møllevangskolen (Figure 17). This allows us to compare the inflows between all four polling places during this hour.

Time	Voters	Devices	Percentage
...
16:00 - 16:10	80	29	36.25%
16:10 - 16:20	74	25	33.78%
16:20 - 16:30	97	27	27.84%
16:30 - 16:40	102	14	13.73%
16:40 - 16:50	108	27	25.00%
16:50 - 17:00	112	33	29.46%
...

Fig. 14. Device Capture Ratio in Holbergskolen Copenhagen

Time	Voters	Devices	Percentage
16:10 - 16:20	90	28	31.11%
16:20 - 16:30	85	33	38.82%
16:30 - 16:40	132	42	31.82%
16:40 - 16:50	122	43	35.25%
16:50 - 17:00	118	47	39.83%
17:00 - 17:10	130	48	36.92%

Fig. 15. Device Capture Ratio in Bellahøj Skole Copenhagen

Time	Voters	Devices	Percentage
16:00 - 16:10	107	56	52.34%
16:10 - 16:20	183	77	42.08%
16:20 - 16:30	145	79	54.48%
16:30 - 16:40	200	84	42.00%
16:40 - 16:50	211	93	44.08%
16:50 - 17:00	193	88	45.60%

Fig. 16. Device Capture Ratio in Islands Brygge Skolen Copenhagen

Time	Voters	Devices	Percentage
16:00 - 16:10	138	49	35.51%
16:10 - 16:20	160	64	40.00%
16:20 - 16:30	170	76	44.71%
16:30 - 16:40	173	76	43.93%
16:40 - 16:50	165	73	44.24%
16:50 - 17:00	201	83	41.29%

Fig. 17. Device Capture Ratio in Møllevangskolen Aarhus

7. FINDINGS

This section describes some preliminary findings – a detailed analysis of the data collected and conclusions to be drawn in left to a different paper.

- (1) The white box method provides a precise and accurate method to collect information for measuring waiting times in polling places. We have observed and recorded data in one polling station in two different ways, using our white box method and a manual counting method. Figure 11 reports our observations. We can see that both data sets (green and light blue) follow (roughly) the same trend, but that the automatic method recorded overall higher waiting times than the manual method. We attribute this difference to that mobile phones are detected earlier by our white box sensors because they record mobile phones while they are approaching the polling station and not just when they arrived. To a lesser extent, it may also be due to the inaccuracy inherent in manual data collection and the low sampling rate.
- (2) Some polling places experienced an increase in waiting times between 16:00 and 18:00. For example, at Islands Brygge Skolen waiting time doubled around 17:00. We suspect that this because many voters voted after work. To alleviate these service bottlenecks, our findings indicate that either more registration desks or more curtains are needed. We can also refine the setup to collect more detailed data, for example, by increasing the number of sensors, or by using location algorithms, for example, trilateralization. It could even be possible to track individual mobile phones or to predict where a line will form. Such extensions are beyond the scope of this paper.
- (3) The white box data can be also be used to analyze the percentages of the voting populations, who have smartphones and this information can be correlated with voter participation. It is also possible to conduct a further statistical analysis, for example by correlating our data with the official voter registration data.
- (4) Our findings also show that every polling station has a characteristic service time. This is the minimal amount time a voter requires to traverse the polling station. In the Islands Brygge Skolen polling station, the minimal service time is around 200 seconds, whereas in Bellahøj Skole, it is closer to 270 seconds. Factors that effect the service time include polling station layout and processing speed.
- (5) In general, the averaging method and Little's method provide very similar results, although the latter appears to be more fine-grained. Especially the plots obtained by Little's method in

the Figures above and the Appendix below show interesting periodic fluctuations, that we will analyze in future work.

- (6) Finally, not surprisingly, all polling places show a peak in waiting times at the beginning of the voting day. We have observed long queues forming outside polling places, before they open. When the voting started, and polling places have been operating for a little while at full capacity, waiting times dropped dramatically.
- (7) In Frederiksbjerg Hallerne (as shown in Appendix B, Figure 37), there is a dip in the graph around 10am, which coincides with the malfunction of sensor No. 7 placed at the entrance at the polling place, as discussed in Figure 3. This means that area covered by our sensors was somewhat smaller during this time, resulting in reduced waiting times in this polling place.

Acknowledgments:

Thanks to Copenhagen municipality for supporting the pilot study, Danmark's Ministry of the Interior and Economic Affairs, and Datatilsynet, the Danish data protection agency for supporting the pilot. In addition, we wish to thank Jens Egholm, Nina Holm-Jensen, Aske Holm-Jensen, Lorena Ronquillo, Christopher Gad, Steffen Dalsgaard and Signe Lund Tovgaard for their help with deploying the white boxes.

REFERENCES

- David Basin, Patrick Schaller, and Michael Schlapfer. 2009. *Applied Information Security*. Springer Berlin Heidelberg, Berlin, Heidelberg. DOI : <http://dx.doi.org/10.1007/978-3-642-24474-2>
- Robert F. Bauer and Benjamin L. Ginsberg. 2014. *The American Voting Experience: Report and Recommendations of the Presidential Commission on Election Administration*. Technical Report January. 01–112 pages. <https://www.supportthevoter.gov/files/2014/01/Amer-Voting-Exper-final-draft-01-09-14-508.pdf>
- Duncan a Buell. 2013. An Analysis of Long Lines in Richland County, South Carolina. *Presented as part of the 2013 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections* 1, 1 (2013), 106–118. <https://www.usenix.org/conference/evt2013/workshop-program/presentation/Buell>
- Danmarks Statistik. 2015. Oversigt over valgkredse i Danmark. (2015). <http://www.statistikbanken.dk/10005>
- John D. C. Little. 1961. A Proof for the Queuing Formula: $L = \lambda W$. (1961). DOI : <http://dx.doi.org/10.1287/opre.9.3.383>
- Motorola Solutions. 2014. *Analysis of IOS 8 MAC Randomization on Locationing*. Technical Report October.

A. DEVICE APPEARANCE

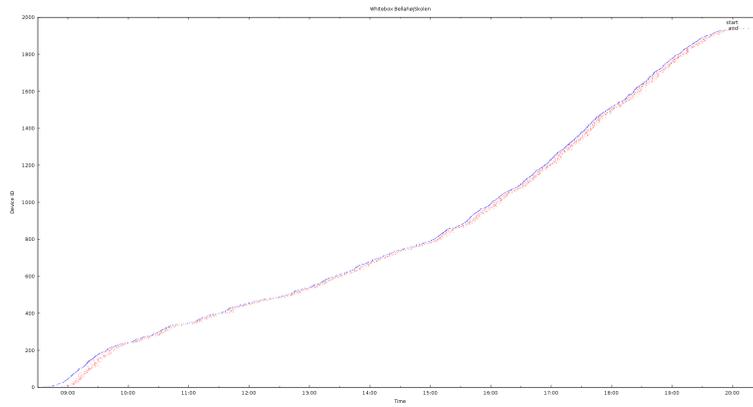


Fig. 18. Device Appearance in Bellahøj Skole Copenhagen

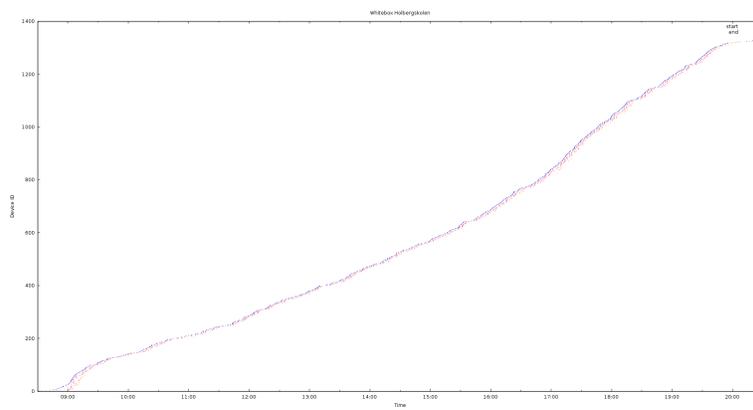


Fig. 19. Device Appearance in Holbergskolen Copenhagen

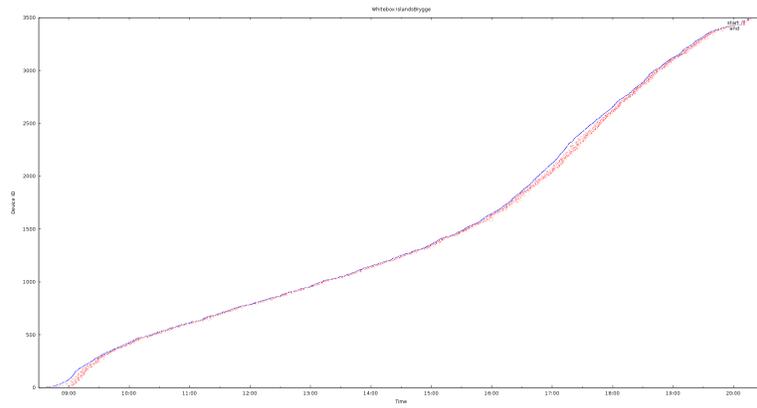


Fig. 20. Device Appearance in Islands Brygge Skolen Copenhagen

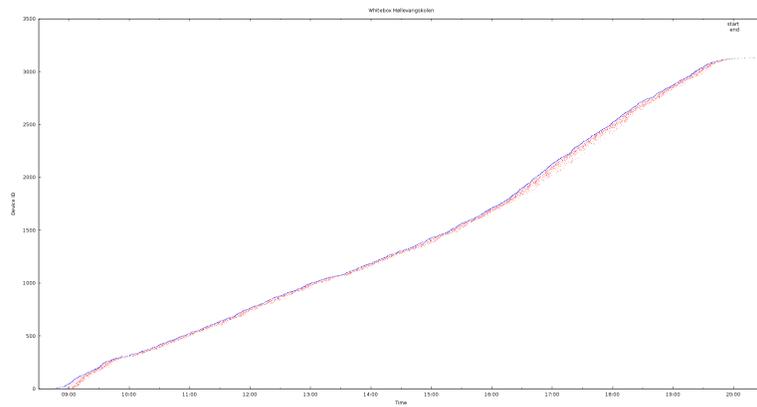


Fig. 21. Device Appearance in Møllevangskolen Aarhus

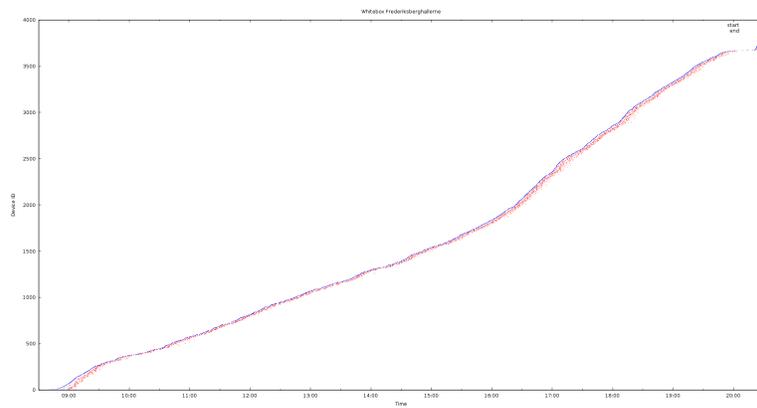


Fig. 22. Device Appearance in Frederiksberg Hallerne Aarhus

B. AVERAGE WAITING TIME VIA AVERAGING METHOD

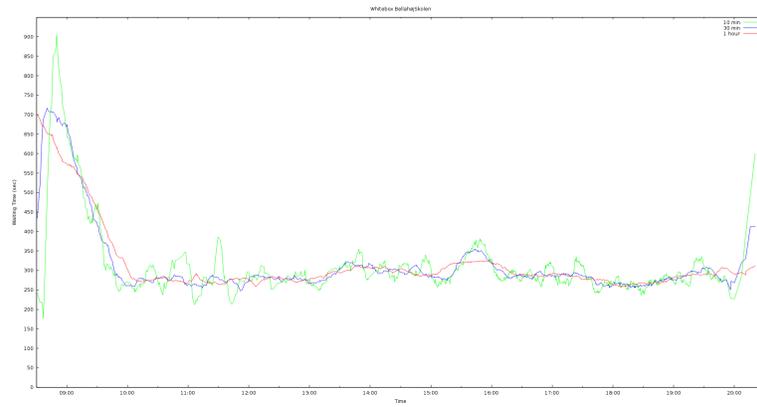


Fig. 23. Average Time (Averaging Method) in Bellahøj Skole Copenhagen

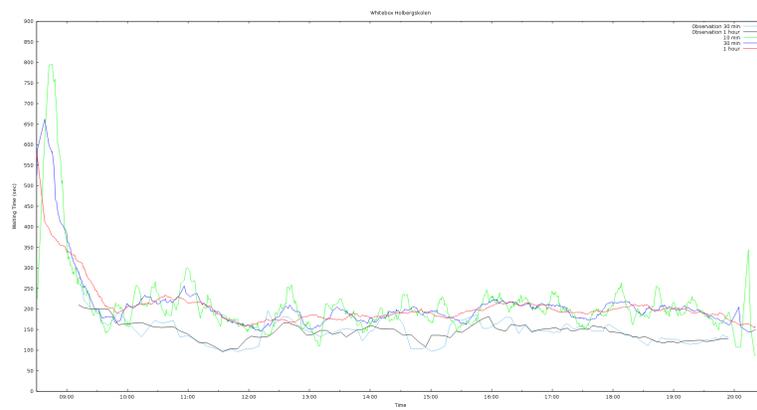


Fig. 24. Average Time (Averaging Method) in Holbergskolen Copenhagen

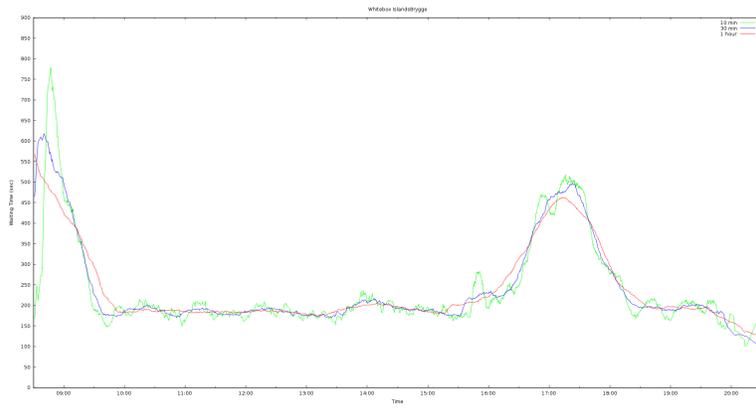


Fig. 25. Average Time (Averaging Method) in Islands Brygge Skolen Copenhagen

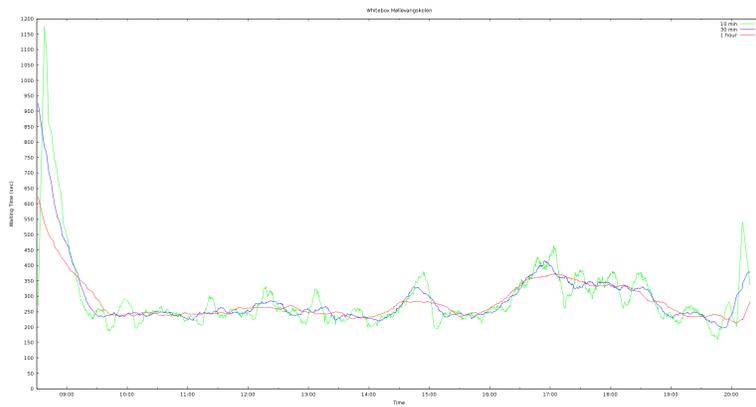


Fig. 26. Average Time (Averaging Method) in Møllevangskolen Aarhus

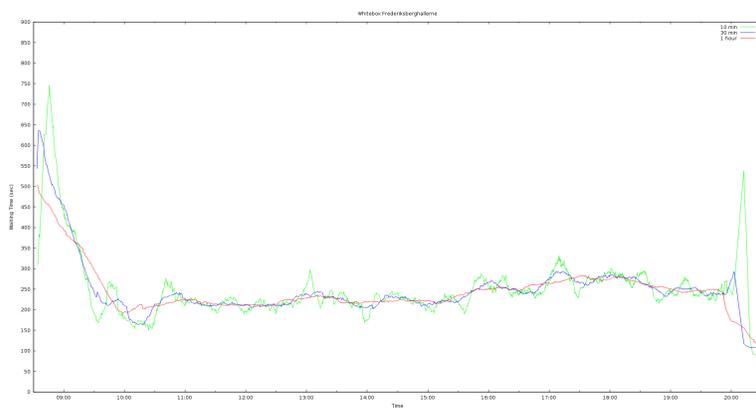


Fig. 27. Average Time (Averaging Method) in Frederiksberg Hallerne Aarhus

C. AVERAGE WAITING TIME VIA LITTLE'S METHOD

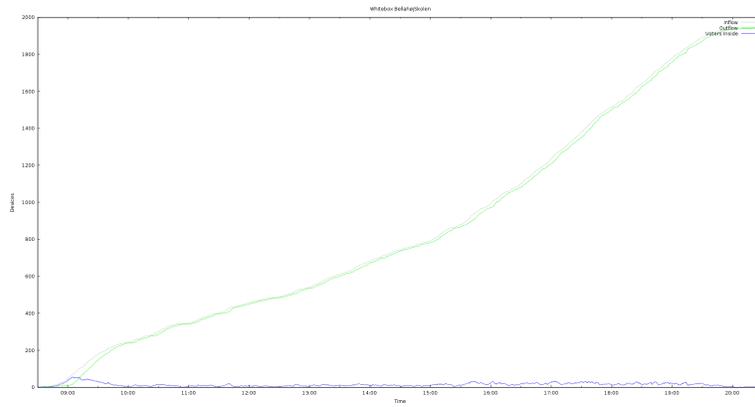


Fig. 28. Device Flow in Bellahøj Skole Copenhagen

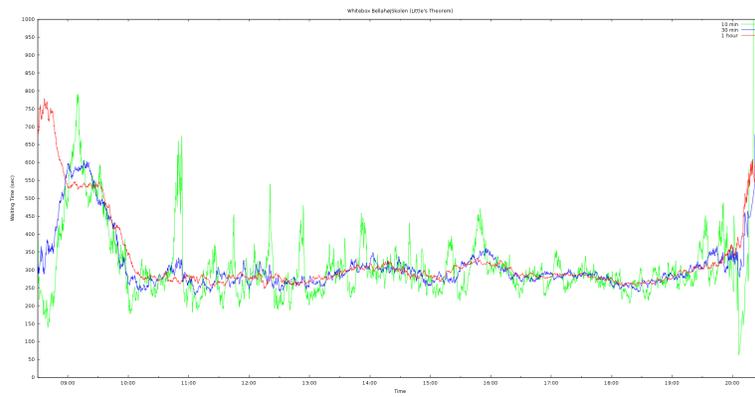


Fig. 29. Average Time (Little's Method) in Bellahøj Skole Copenhagen

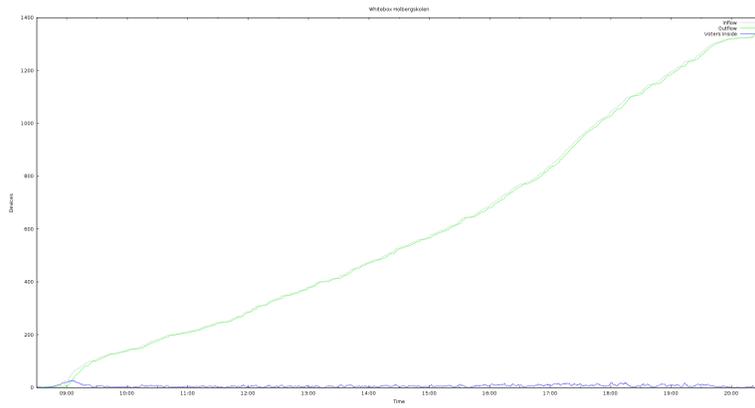


Fig. 30. Device Flow in Holbergskolen Copenhagen

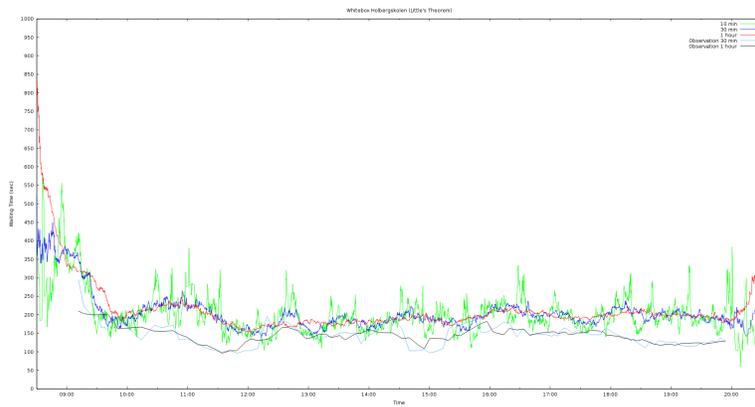


Fig. 31. Average Time (Little's Method) in Holbergskolen Copenhagen

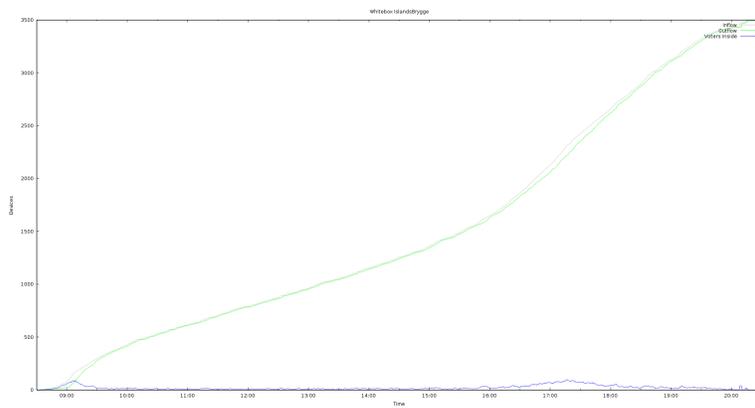


Fig. 32. Device Flow in Islands Brygge Skolen Copenhagen

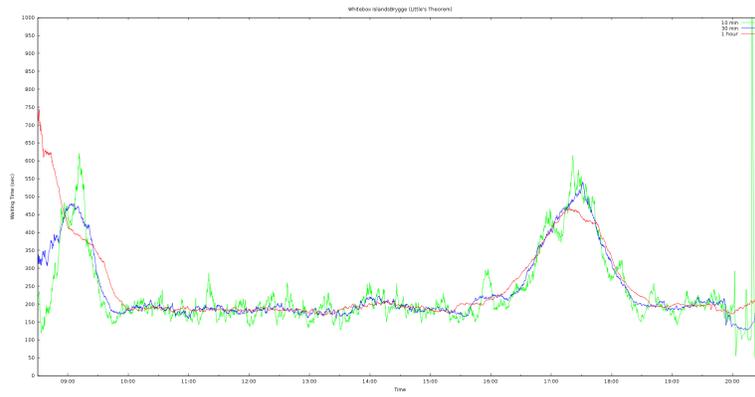


Fig. 33. Average Time (Little's Method) in Islands Brygge Skolen Copenhagen

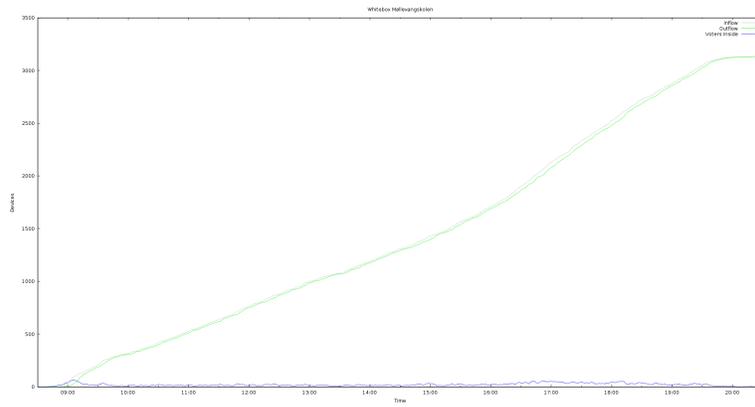


Fig. 34. Device Flow in Møllevangskolen Aarhus

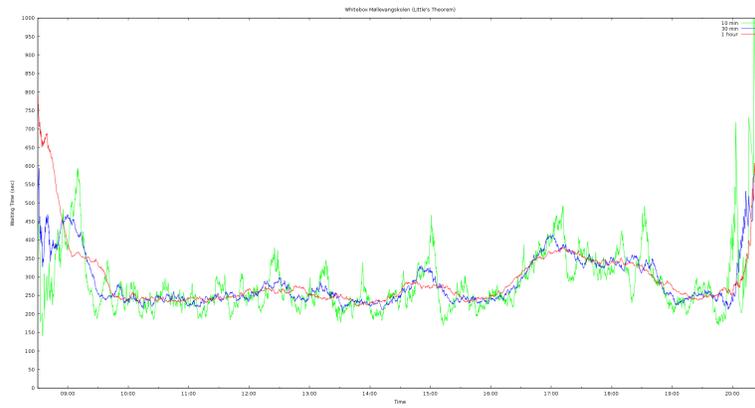


Fig. 35. Average Time (Little's Method) in Møllevangskolen Aarhus

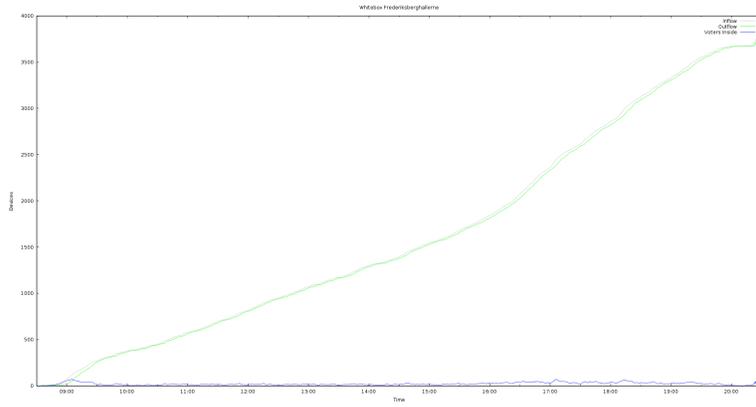


Fig. 36. Device Flow in Frederiksbjerg Hallerne Aarhus

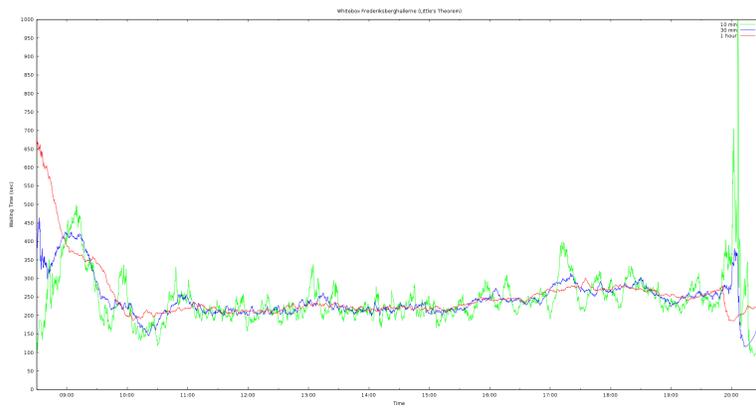


Fig. 37. Average Time (Little's Method) in Frederiksbjerg Hallerne Aarhus