RF-based Passive Crowd Density Estimation in Realistic Large-Scale Environments

Stijn Denis, Ben Bellekens, *Member, IEEE* Rafael Berkvens, *Member, IEEE* and Maarten Weyn, *Member, IEEE*

***Abstract*—Automatic crowd estimation can be a highly useful tool for organizers of large-scale events, particularly in the context of crowd safety management. In recent years, the use of an RF-based passive sensing approach to design this type of system has received increasing research interest. However, the vast majority of systems described in literature were experimentally deployed in environments which contained at most a few tens of human individuals. In this paper, we describe our own passive crowd estimation solution which we deployed in a variety of real-life, large-scale setups over the course of two years. We show that the average RSS-attenuation measured within an RF-network we install in an environment provides a reliable indication regarding the size of a crowd which can contain thousands of individuals. This is validated through two of our setups for which we had access to ground truth data. In both cases, the correlation coefficient between the measured attenuation and the actual crowd size exceeds 0.95. Additionally, the use of linear regression to create an actual crowd estimation model led to a median estimation error of approximately 191 human individuals for an environment in which a crowd containing more than 3000 people was present.**

***Index Terms*—Crowd Estimation, Crowd Analytics, Crowd Counting, Footfall Analytics**

I. Introduction

OVER the past few years, the development of automatic crowd estimation systems has become a popular topic of research. These systems are capable of estimating the number of people present in a crowd - an important metric in a multitude of contexts. Traffic control, popularity assessments at commercial events and crowd safety management are all examples of situations in which automatic crowd estimation can be used. Crowd safety management in particular has been noted in literature as a core application [1], given that many tragic cases exist of fatalities caused by mass panic breaking out at large-scale events [2]. A more accurate understanding of crowd sizes could potentially aid event organizers in managing the crowd in order to prevent such situations from occurring. The use of a visual camera-based approach is a common solution to this issue. It has several downsides, however. Lighting conditions need to be relatively stable for the accuracy of these systems to remain consistent and the targets we wish to count cannot be obscured by smoke or darkness. Furthermore, these kind of approaches tend to require quite a lot of computing power which may not always be readily available. Finally, the use of vision-based technologies will always have the potential to cause privacy-related issues. Another possible solution is to make use of the impact the crowd has on RF-waves in a wireless sensor network (WSN). These kind of approaches fall under the umbrella of 'passive' or 'sensorless' sensing, in which radio transceivers are not only used for communication, but also act as sensors. This concept has become a popular topic of research over the last decade and systems utilizing this principle have been developed for device-free localization (DFL) [3], activity recognition [4], fall detection [5] and gesture recognition [6].

Manuscript received xx, xxxx; revised xx, xxxx.

Crowd estimation as a research subject can be considered to belong to the detection aspect of device-free localization [7] and quite a few RF-based systems have been developed which focus on this facet [8]–[10]. Many of these approaches solely utilize signal strength measurements, but in recent years several techniques have appeared which incorporate Channel State Information (CSI) as well [11].

Unfortunately, the vast majority of the systems described in literature were installed in relatively limited environments in which the crowds were no larger than a few tens of people. While the potential benefits of using these types of systems in crowd management scenarios are regularly discussed [8], [10], almost no experiments have been performed thus far in realistic, large-scale environments containing thousands of human individuals. To the best of our knowledge, our own RF-based setup at a Belgian music festival in 2017 which we described in [12], comprises the sole exception. While this experiment was still very much preliminary in nature, the results did indicate the feasibility of using the sensorless sensing concept for large-scale crowd estimation and served as a springboard for our continuing research.

Since then, we have installed our large-scale crowd estimation system at a variety of different real-life events: Tomorrowland 2018, Music for Life 2018, Sound of Science 2019 and Tomorrowland 2019. In each case, the results of these experiments were shared with the event organisers afterwards and it was shown that even comparatively simple analyses of the collected data could provide highly interesting crowd-related insights. As a result, during one of our most recent setups, we were asked to provide real-time crowd data to the control center of the Tomorrowland 2019 music festival.

In this paper, we will provide an overview of the designs, implementations and results of the crowd density estimation systems we have installed in real, commercial environments over the past two years. It should be noted that we use the terms crowd density, crowd estimation and crowd counting interchangeably. At the moment, we are only interested in the number of people present within a certain predefined environment and research questions regarding the crowd shape are considered to be out of scope.

We believe our system to be particularly interesting in

an IoT-context. For most of our early experiments, it was sufficient to collect data in a certain environment and analyze the results afterwards. Our more recent (real-time) implementations have increasingly shown us the need for our setups to be integrated within already existing, real-time ecosystems in which other types of event-related data are also being collected (e.g. number of cashless payments per minute, amount of people who had their entrance bracelet scanned to gain access to VIP areas, ...). This will provide excellent opportunities for data fusion which could lead to additional crowd insights for the event organizers. Furthermore, we believe that the growth of the Internet of Things will have a significant impact on the development of sensorless sensing approaches, as the increasing presence of RF-signals will enable more widespread deployment of passive techniques which make use of these signals. An LTE-based crowd estimation system proposed by Di Domenico et al. [10] is a good example of this concept which makes use of already available network infrastructure. Finally, it must be noted that the general concept of sensorless sensing in an IoT context has previously been discussed in literature, as exemplified by the work performed by Kianoush et al. in [13]. They proposed a cloud-IoT platform for passive RF-sensing based applications and evaluated its performance through two concrete use cases related to assisted living and driver behaviour recognition.

The remainder of this paper is structured as follows. In the next section, we will discuss the architecture of the system and the manner in which it evolved throughout the past two years. A conceptual overview of the entire system will be provided, as well as a more in-depth explanation of the components which make up the whole: hardware, communication schedule, backend and real-time visualization. Section III provides an overview of the performed experiments in different types of environments and the crowd estimation results that were obtained therefrom. These results open up a variety of future research paths, which we describe in Section IV. Finally, in Section V, a conclusion is presented.

II. System Architecture

To enable robust crowd density monitoring, a specific architecture was created to provide a production-ready system for large-scale environments that facilitates a fast and scalable deployment. In order to satisfy these requirements, the architecture is based on separate purpose-based programs that process the data via a message broker through different topics. The programming language that we used to develop the Backend system is Python. Moreover, to make the system scalable and robust, all programs are deployed in a container-based cluster. This enables fast deployments and auto-restarts when a program crashes. Furthermore, a management service was developed to interact with the different programs to allow for real-time reconfiguration. From a top-level perspective, the system primarily consists of a backend system that is capable of processing and storing RF-data in real-time so that it can be fetched by a visualization platform (e.g. Grafana). To capture this data, a set of transceivers is installed around and/or within the environment in which the crowd is present. These transceivers are able to communicate with each other and with a gateway, which passes every measurement towards the backend for further processing. In Figure 1, the system architecture is shown schematically.

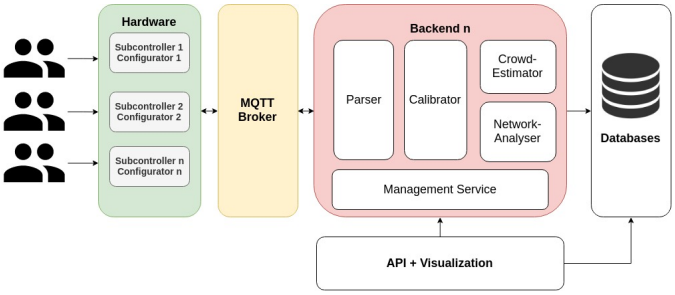


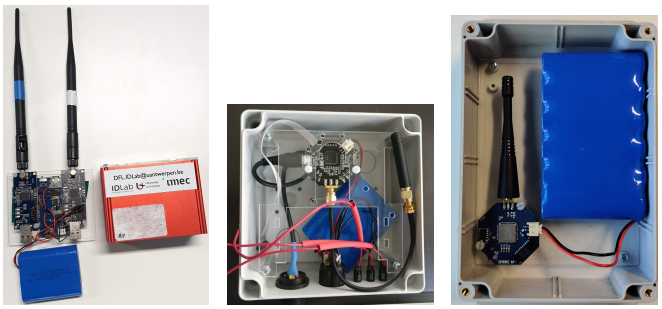
Fig. 1. Schematic overview of our passive crowd estimation system architecture.

The current system architecture is the result of multiple years of experience and testing in critical production environments. Changes were primarily made for reasons related to scalability and robustness. In this context, all Python programs were initially executed directly on the scheduler of the host-CPU without any fail-over system or a central storage mechanism for logging. To ameliorate issues arising from these early setups, a central container-based backend was configured that could restart itself when an exception was triggered. Furthermore, container-based systems can have centralized logging and inter-network connectivity capabilities and can be deployed quickly, which resulted in a more stable system. Finally, the backend container itself was divided into multiple containers so that every container was created for a single task, which brought more scalability and robustness because each task could be tested and deployed separately.

A. Node Hardware

As outlined in the previous subsection, the RF-data collection system consists of a gateway and a series of nodes. Over the course of the past two years during which the experiments were performed, multiple types of hardware were used to fulfill these roles.

Figure 2 shows the major three types of physical nodes that were deployed in our setups. The first type consisted of an aluminium casing which contained a 6600 mAh battery,



(a) Type I (b) Type II (c) Type III

Fig. 2. Illustration of the physical nodes.

a LiPo Rider Power module and two self-developed EZR-USB transceivers with RF-frontends that were optimized for respectively 433 MHz and 868 MHz communication. Two sides of the casing were open, of which one was covered by a transparent plastic plate. Antennas could be directly connected to the EZR-USB modules through holes in this plate. Because the antennas were on the outside of the node - and therefore clearly visible - we were usually required by the event organizers to install this type of node in locations which were out-of-sight for the public (e.g. under wooden bars where only staff members could see them). As previously discussed in [12], this was the first type of node we created and it was primarily used for measurements within indoor environments: the Freedom stage at the 2017, 2018 and 2019 editions of Tomorrowland and the Flame environment at De Warmste Week in 2018. The second node type had a casing which was entirely closed and made of plastic. It contained a 6600 mAh battery and two octa-mini Octa-Connect modules containing an EZR microcontroller. Antennas were attached to the modules through a connector. In contrast to the previous node type, these antennas were on the inside of the casing, which meant that it tended to be less problematic for these nodes to be publicly visible. Furthermore, the outside of the casing contained a power switch, a USB-port which could be used to charge the battery and a green, yellow and red LED which were steerable by the octa-mini modules to indicate node status. In practice, however, these LEDs were not used in actual setups. Considerable variation existed within this node type; some nodes contained only an 868 MHz module, while in others an additional 2.4 GHz Bluetooth module was present which could be used for other (smaller-scale) sensorless sensing projects. Nodes of this type were used in a variety of different setups and they were generally considered to be our standard node type. Finally, the third type of node consisted of a plastic casing containing an 15000 mAh battery and a single 868 MHz octa-mini module which contained a Murata microcontroller and was directly attached to an antenna. This type was primarily used in an important part of our setup at Tomorrowland 2019 in which nodes were deployed at several critical locations alongside a large bridge which was a main thoroughfare of the festival.

B. Communication Schedule

The general manner in which the node network operates, did not significantly change between the experimental measurements described in [12] and our more recent experiments. The DASH7 Alliance Protocol (D7AP) [14], which is meant to be primarily used for mid-range IoT use cases, still forms the basis of all communication that occurs within the network. It should be noted that while DASH7-communication networks generally use a star or tree topology, this is not the case for our system as each node communicates with every other node. In the next few paragraphs, we will provide a more detailed explanation of the communication schedule.

From a functional point of view, each network contains three different types of nodes: regular nodes, a controller node (controller) and a configuration node ('configurator'). Communication occurs in a time slotted manner in which each regular node in turn broadcasts a data packet towards all other nodes. The RSS-value with which a packet is received is saved by each node in an internal list. When the time for a node to broadcast has arrived, the data packet it transmits will contain the values of this list. After the transmission is complete, the internal list is then entirely cleared. Meanwhile, the controller node listens to all communication within the network and passes the received RSS-values on to a computing unit to which it is connected (a laptop, a NUC, a Raspberry Pi, ...). This means that each transmission serves two purposes: measuring new RSS-values and communicating earlier RSS-values to the backend through the controller.

Additionally, the controller is also responsible for synchronizing the node communication and it instructs the nodes to start or stop communication cycles. It can be directed to do so through sending certain commands over a UART connection. Commands can also be sent to a controller through specific DASH7-messages, but this feature has not been actively used in practical setups thus far. Other commands include a reset command for the controller, a reset command for the entire network (which leads to the controller broadcasting an instruction causing all nodes to reset themselves, after which the controller resets itself as well) and an update configuration command. This update configuration command causes the controller to broadcast a message which, when received by the nodes, causes them to enter their setup mode. In this mode, nodes repeatedly broadcast configuration request messages until they receive a valid response from a configuration node. This setup mode is also automatically entered by every node - including the controller - directly after booting.

The configuration node contains a variety of different network parameters and is continuously listening for configuration request messages. Aside from determining the order in which nodes transmit during a communication cycle, the configuration node will also transmit the size of the network and an entire DASH7 access profile, which defines many communication-related parameters such as the frequency channel(s) on which communication should occur. Communication between a node and the configurator always occurs using a fixed, hard coded access profile which is entirely separate from the profile that is used for all other communication within the network.

An important advantage of working with a separate configurator is the fact that when one wishes to change one or more network parameters (e.g. adding/removing nodes or changing the frequency channel), only the firmware of the configurator needs to be reflashed. Additionally, if a node encounters an error, crashes and resets itself, it will enter its setup mode and once again start sending configuration requests, without causing any interference with the currently ongoing cycle. Once it has obtained the necessary parameters, it can simply join the next cycle.

C. Backend

As explained previously, the backend is responsible for processing and storing the RF-data of a single environment in

real-time. In order to handle the RF-data of each environment for real-time processing, each gateway will pass the raw measurements that are received from a controller to a central message broker. In most of the cases we deployed each back-end on a central processing unit or private cloud infrastructure together with the message broker and the databases to reduce latency and increase robustness. Within the structure of our backend, two data-flows are described: the estimation flow and the calibration flow.

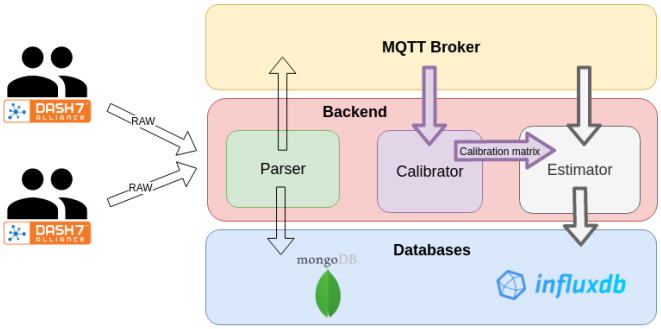


Fig. 3. Data-flows within our system architecture: purple arrows indicate the calibration data-flow and grey arrows indicate estimation data-flows on top of the generic data-flow.

During the first step of the estimation flow, the parser translates the raw DASH7-data into meaningful structured data, which describes the RSS-values, node-id, cycle-id, and controller hardware address. During the second step, the estimator will buffer each data packet for a specific period of time. For our deployments, a delay of either 5 or 10 seconds was used to increase the stability of the estimation and reduce the sensitivity of possible outliers. Next, the estimator will process the parsed data by estimating the average of the differences of each buffered received signal strength relative to a calibration file, which is a matrix of the average RSS-values measured when the environment was (largely) unoccupied. The calibration flow is responsible for creating this square matrix, which has a size equal to the total number of nodes. In addition to processing the data, both data-flows will store the data in databases. Each DASH7 data-message is stored in a MongoDB collection, while every average difference relative to the calibration matrix or estimation is stored in a time-series InfluxDB database. Furthermore, the number of received messages per communication cycle - which we consider to be indicative of the RF-network reliability - is calculated by the network-analyser and stored in the InfluxDB for visualization and further analysis. Finally, we have also developed a data-labeler which annotates the parsed data based on a specific label. Such a label can be a simple user-generated string that describes the environment or event during a specific range of time-stamps or a single time-stamp. Annotations can be very valuable for post-processing to aid us in interpreting the data. Furthermore, they can be used for future prediction purposes when more data sources will be integrated within the system. In Figure 4, the dashboard of the Main Comfort environment at Tomorrowland 2019 is shown.

In order to provide such a dashboard, the visualization platform Grafana was deployed on the central processing unit

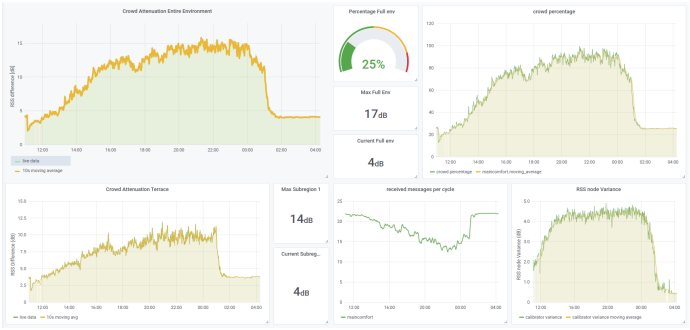


Fig. 4. Illustration of our crowd density estimation dashboard.

or on a private cloud infrastructure. Grafana will directly interact with the InfluxDB by querying the database on specific intervals. As shown in the figure, different graphs, values, and a gauge are displayed to show the current estimated crowd density values within the time-interval that was configured in the dashboard. These values indicate the occupancy percentage, average RSS-difference of the entire environment and the average RSS-difference of the terrace subregion. Occupancy percentage was determined based on the maximum RSS-difference value that had been observed within this environment. Moreover, the results of the network-analyser are displayed, as well as the variance of the RSS-differences for all the communication links. This last parameter can be used for future auto-calibration purposes, which will be explained in more detail in Section IV.

III. Crowd Estimation

A. Experiment Overview

As mentioned in the introduction, in this paper we discuss the experiments we have performed at 4 different events over a period of 2 years: Tomorrowland 2018, De Warmste Week 2018, Sound of Science 2019 and Tomorrowland 2019.

1) Tomorrowland 2018: At Tomorrowland 2018, we repeated our experiment from [12] at the Freedom stage. Additionally, we were allowed to install a system at the Main Comfort location which provided a clear view of the Main stage. 54 nodes were installed at the edges of an environment of approximately 1570 m^2. The Main Comfort could only be entered by visitors who had VIP access. As a result, they had to have their bracelets scanned when entering and leaving. The organizers made use of this scanning principle to approximate the number of people currently present. We were allowed access to the resulting data and it provided us with a comparatively decent ground truth. However, it was by no means an exact, accurate count. This was primarily due to two reasons. First of all, not everybody was scanned when entering or exiting the Main Comfort. Crew and staff members had to manually show their separate type of bracelet and visitors leaving the environment at the very end of a festival day generally were not scanned either. Second, the Main Comfort environment was present at the bottom floor of a 3-floor construction, with the more exclusive B2B and Skybox environments being housed in respectively the second and third floors. Visitors with the required access rights could

enter these environments through the Main Comfort and had to have their bracelets scanned a second time. We did not have access to the data from this second set of scanners, which meant that our ground truth data also (partially) included people who were present on the second or third floor, despite our system only being installed at the ground floor. Nevertheless, it provided us with our first (and as of present, only) decent approximation for an environment capable of containing thousands of human individuals. In Section III we will illustrate our crowd estimation approach primarily based on results from this experiment.

2) Music for Life 2018: Music for Life is a yearly charity event organized by the radio station Studio Brussel which is owned by the Flemish public broadcaster VRT. As part of the De Warmste Week format of the event, the organizers set up an outdoor radio studio at a recreational domain from which they broadcast during the week leading up to Christmas. This location can be visited by the public and during the evening hours, several bands perform live music. Additionally, tickets can be bought online for longer concerts which take place in the nearby Flame tent. For our experiments, we installed 45 nodes in the main Radio environment (approximately 9000 m^2) and 44 nodes around the edges of the Flame environment (approximately 625 m^2). An interesting aspect of the Radio environment was the fact that it contained a series of metallic towers to which spotlights were attached. These towers were surrounded by wooden pallets to which we could attach our nodes. This meant that this was our first test location where we could install nodes within the environment and not just around the edges. A challenging characteristic of both setups was the fact that the controllers for both node networks were present in a single physical gateway. This gateway was installed on top of a stationary truck which was parked in between the two environments. The large amount of metallic materials used in the construction of the Flame tent and the fact that the distances between the gateway and the nodes could go up to 100 meters, made this entire experiment an interesting connectivity test in a difficult environment.

3) Sound of Science 2019: The Sound of Science festival is a yearly Belgian event which aims to increase interest in science within the broader public. It does so through a variety of interactive demonstrations, performances and other scientifically minded activities. As part of the 2019 edition, we performed a controlled large-scale crowd estimation experiment. 44 nodes were deployed in a tent environment called the 'Nucleus Tent' which had a size of approximately 580 m^2. Due to the presence of several support poles, 8 nodes could be attached to them and therefore placed within the environment, while the rest were installed around the edges. Additionally, the nodes were placed at a height of approximately 1.20 meters, significantly higher than what had previously been the case. A major advantage of this experiment was its controlled nature. Unlike our previous measurement scenarios, this time we were able to order a volunteering crowd (with a maximum size of 247 human individuals) to sit down, stand up and have specific parts of the crowd leave the environment at predetermined moments in time. In doing so, we were able to obtain an extremely accurate ground truth.

4) Tomorrowland 2019: For Tomorrowland 2019, we tripled the number of RF sensing networks we deployed when compared to the previous edition. First of all, we repeated our experiments in the Freedom and Main Comfort environments, albeit with a significantly lower number of nodes (35 and 19 respectively). Additionally, 20 nodes were installed in the B2B environment situated directly above the Main Comfort and 18 nodes were installed within the final floor of the building - the Skybox environment - as well. Finally, a system was also deployed at several critical locations that were part of a large wooden bridge which was a main thoroughfare for the festival. Due to the large size of the bridge and the resulting distances between some of the nodes, the decision was made to split the deployment in this environment into 2 RF-networks. Tomorrowland 2019 marked the first time in which the full system as described in Section II was implemented. Additionally, this was the first edition of the Tomorrowland festival in which we at their request - directly provided our resulting real-time data to the control center of the event organizers. Unfortunately, at present we have not yet received ground truth data for the Main Comfort, B2B and Skybox environments and the corresponding analyses have therefore not yet taken place.

B. Calibration

The results of the experiment performed in our earlier study indicated that the average attenuation experienced by communication links within an RF sensing network, is a useful parameter for determining crowd density. These attenuation values can be determined by calculating the differences between current RSS-measurements and RSS-measurements which were performed when the environment was (largely) unoccupied by human individuals. The result can be negative - which means that the communication links within the environment are on average experiencing positive interference - but this only happens when occupancy is extremely low. In Figure 5, a graph is presented which shows the evolution of this average attenuation throughout a single day at the Freedom Stage of Tomorrowland 2018. Each data point was calculated based on communication that occurred within a 10 second interval.

This means that it is necessary to delineate a calibration period in time during which the environment is both static and unoccupied in order to perform calibration measurements. Furthermore, during medium-to-long term setups in which the position of generally static objects in the environment can change over the course of the measurements (e.g. slightly changed positions of the tables in the Main Comfort environment from one day to the next), these measurements need to be regularly updated. Early analyses indicate that it is possible to automatically determine the time intervals which are most suitable for these measurements to occur, without needing constant visual verification that the environment is truly empty. We observed that by analyzing the mean variance experienced by communication links, periods of time in which the environment was largely static and which were therefore prime candidates to be used as calibration periods

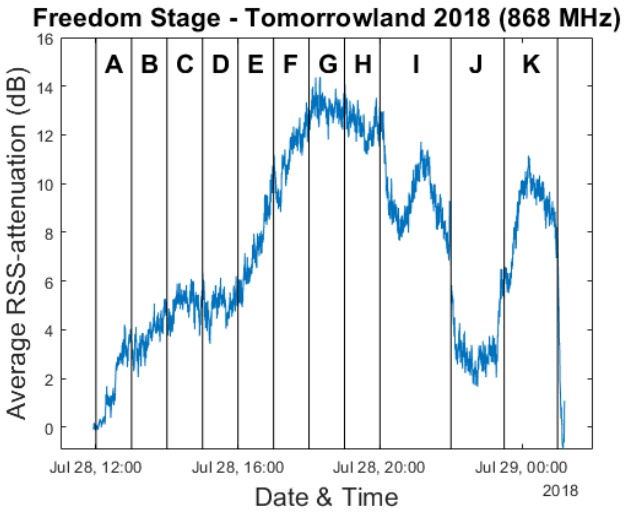


Fig. 5. Average attenuation within our installed RF-network (Freedom Stage at Tomorrowland 2018). The letters separated by vertical lines indicate different artists performing on stage during the day.

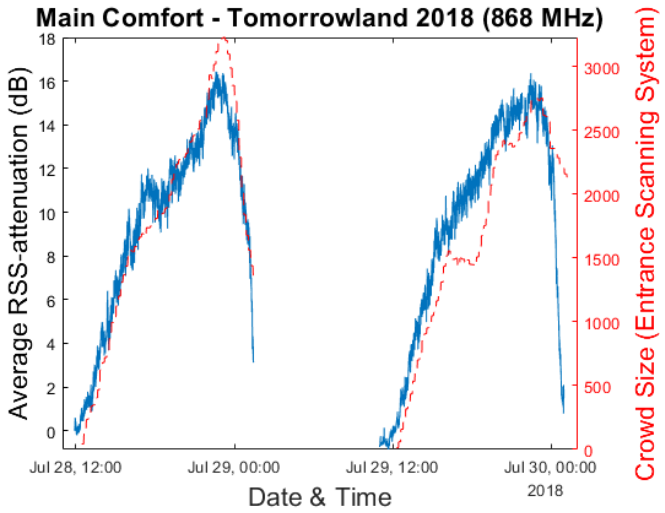


Fig. 6. Average attenuation within our installed RF-network and crowd sizes as determined by the entrance scanning system (Main Comfort at Tomorrowland 2018).

- could be detected. This observation forms the basis for an autocalibration system which we are currently developing. We will explain this in more detail in Section IV.

C. Estimating Crowd Sizes

We can determine the average attenuation measured within the RF-network we deployed within an environment and we do have indications that this value provides information regarding the size of the crowd that is currently present. The next step consists of translating the average attenuation into an actual crowd count. As previously mentioned, the initial methodology we used in [12] consisted of training a probabilistic neural network to categorize a pair of 433 MHz and 868 MHz-average attenuation values according to 1 of 6 crowd size categories. During subsequent analysis we discovered that the added value of combining multiple frequencies was negligible in this context and therefore, we now generally consider the use of 868 MHz-data to be sufficient. Our first approach which resulted in an actual numerical estimate of people was developed based on the data obtained at the Main Comfort environment at Tomorrowland 2018, a setup for which we had decent numerical ground truth data. Figure 6 compares the graphs of the average attenuation and the number of people present according to the scanning system during the final two days of the festival. Strong correlations can clearly be observed. We calculated the Pearson correlation coefficient between these two variables and a value of 0.97 was obtained. The corresponding scatter diagram is shown in Figure 7. It should be noted that our calculations only took into account data from the official start of a festival day (12:00) until its official end (0:00 or 01:00). As previously mentioned, people leaving the environment after the end of a festival day were generally not scanned, which can be seen by the extremely high crowd counts in Figure 6 at the end of each day.

As a result of discovering the existence of such a strong linear correlation, we investigated the accuracy of a crowd

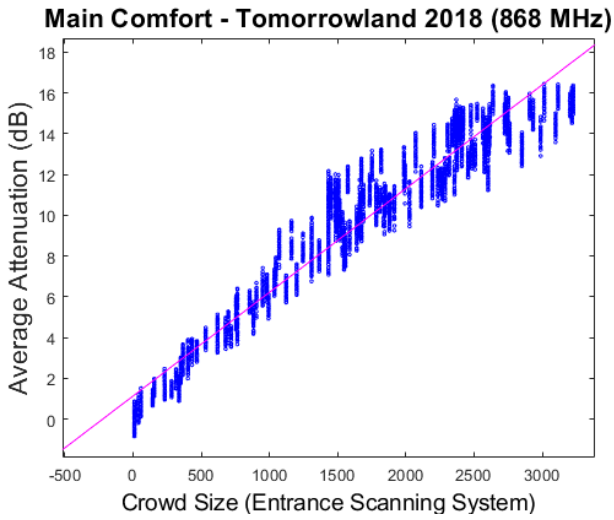


Fig. 7. Scatter diagram showing the relationship between the average attenuation within our installed RF-network and the crowd size as determined by the entrance scanning system (Main Comfort at Tomorrowland 2018).

estimator based on simple linear regression. To determine the linear equation which describes the relationship between the average attenuation and the number of people present, three different sets of training data to fit a line through were used. In each case, the training data was selected from a single festival day and all data from the other day for which we had a ground truth was used for evaluation. Each day was used in turn for training and evaluation and the resulting errors were combined. If a crowd count estimation was a negative number, it was automatically set to 0. The first set contained all data from the entire training day, the second set consisted of data collected during the first two hours of the training day and the final set contained only two data points: the origin and the maximum average attenuation measured during the training day (and the corresponding crowd size). These three datasets were chosen specifically due to practical considerations regarding

an actual commercial crowd estimation system. Depending on the use case, vastly different amounts of training data could be collected in a commercial deployment. For some applications, it would be possible for no training opportunities to even exist at all, which we discuss in more detail in Section IV. The three cumulative distribution function plots of the resulting crowd estimation accuracies are shown in Figure 8. As can be seen in these graphs, comparatively high crowd estimation accuracies were obtained using only a simple linear regression-based approach, with median errors of respectively 229, 286 and 191 individuals for the first, second and third training approaches. Interestingly, the simplest training methodology led to the highest median estimation accuracy. While it should once again be stressed that our ground truth for this setup was not perfect, these results were nevertheless impressive and based on discussions with the festival organizers at Tomorrowland seemed to indicate the viability of the system for commercial applications.

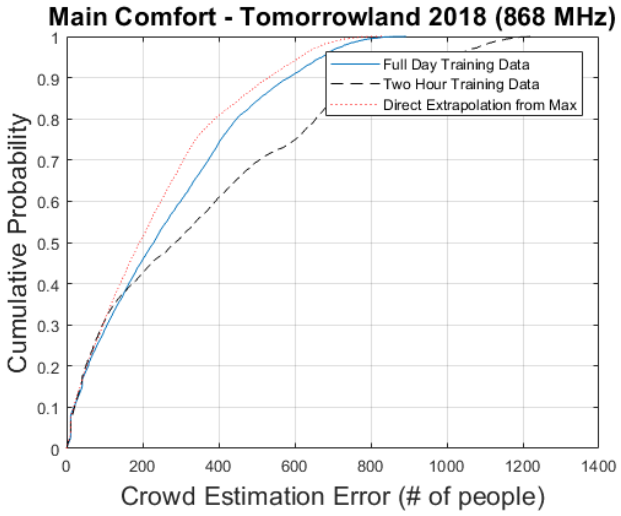


Fig. 8. Cumulative distribution function plots regarding the estimation errors when using three different training approaches for a linear-regression based model (Main Comfort at Tomorrowland 2018).

In order to validate this approach and determine whether the observed strong linear correlations existed for other setups in different environments, a similar analysis was performed on the data collected during the Sound of Science 2018 experiment. While the maximum number of people present was much lower for this environment (247 people instead of over 3000) and the experiment took place over a much shorter amount of time, our ground truth data was nearly exact. We calculated the correlation between the crowd count using only the data which we obtained while the crowd was static and people were not currently leaving the environment, which meant that we knew nearly exactly how many people were present. The resulting comparative graph and scatter diagram can be seen in Figures 9 and 10. With a Pearson correlation coefficient of 0.99, the expected linear correlation still held true. Due to the short duration of this experiment, each average attenuation data point was calculated based on all measurements performed during a 5-second interval.

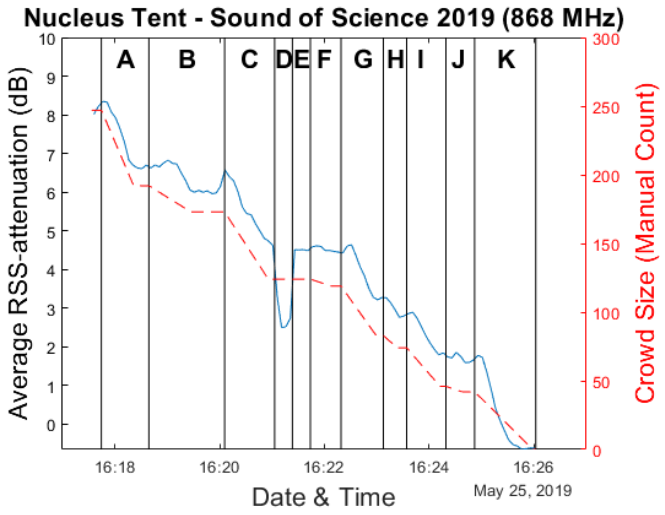


Fig. 9. Average attenuation within our installed RF-network (Nucleus Tent at Sound of Science 2019). The letters separated by vertical lines indicate different experiment phases. An experiment phase generally consisted of a group of individuals leaving the environment. D and E were the exceptions to this, as during these phases the remaining crowd members were asked to respectively sit down and stand up again.

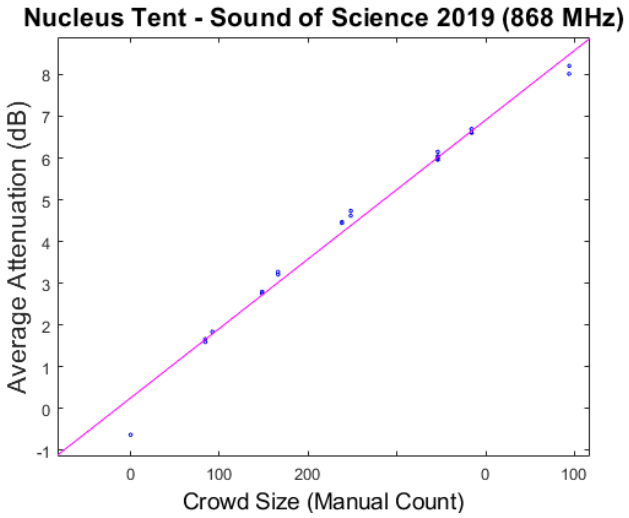


Fig. 10. Scatter diagram showing the relationship between our installed RF-network and the actual crowd size (Nucleus Tent at Sound of Science 2019). In this diagram, only data obtained while the crowd was static (which meant that the exact crowd count was known) is used.

Currently, we do not have numerical ground truth data for any other experimental setup. Generalizing the assumption that a linear relationship exists between the average attenuation and the number of people present, however, did generally lead to a realistic view of changing crowd sizes for our other setups. This was confirmed by manual analysis of camera images (when available) and the expert opinion of event organizers.

IV. Future Work

In the previous sections, we described the inner workings of our system and showed that a large-scale crowd estimator which used a limited amount of training data in combination with simple linear regression can obtain sur-

prisingly accurate results. This has opened up a lot of new avenues of research, some which we have already begun to explore. First of all, an important research direction which we are currently pursuing, is an investigation regarding the relationship between the number of nodes we install in an environment and the resulting crowd estimation accuracy. Up until now, our general approach when installing nodes has been to place as many as is feasible given environmental and organizational constraints. Early analyses of our experiments in which RSS-measurements from many nodes are entirely discarded do seem to indicate that vastly lower amounts of nodes could have been used. This needs to be studied in more detail and optimized numbers of nodes for different types of environments need to be determined. Next, we have the development of the previously mentioned autocalibration system. Automatically detecting when the environment is close to entirely static and then selecting and updating calibration data, would be highly beneficial from a practical point of view. Furthermore, it would enable a much more consistent approach regarding the selection of calibration data than had been the case thus far. An additional avenue we wish to pursue, is the development of a training-less crowd estimation approach. In its current form, the system can be installed within an environment and provide an automatic overview of relative crowd size changes over time with the assumption being made that the relationship between the measured average attenuation within the network and the actual crowd size is linear. Based on our discussions with event organizers, this level of crowd information was already deemed to be highly useful. If some level of training is allowed, the system can even provide a comparatively accurate crowd count. However, providing an actual crowd count without having to perform training measurements in the same environment in which the system was installed, is currently not possible. The development of this capability is a major research track in and of itself and will be one of our primary focuses in the near future. Finally, we are of the opinion that an in-depth study regarding the potential added value of combining multiple data-sources with our own approach is required. One specific tool which safety and security managers at large events or cities are highly interested in, is a forecasting model that is able to predict the crowd density within a short time period. We believe that, in order to obtain such a model, many different data-sources (e.g. weather data, event line-up data, current staff occupancy, ...) need to be combined in an intelligent manner.

V. Conclusion

We have described the full architecture of a passive crowd estimation system which operates based on the physical influence of a human crowd on RF-communication within the environment. In contrast to similar passive RF-based sensing approaches, this system was installed in a variety of real-life setups in large-scale, commercial festival environments which could contain thousands of individuals. The average attenuation within the RF-network, which was measured by comparing current RSS-measurements to an earlier set of calibration measurements when the environment was unoccupied, was shown to be a very strong indicator of the current crowd size. In two environments for which we had access to ground truth data, the correlation coefficients between the variables representing this average attenuation (in dB) and the actual number of people present, were equal to respectively 0.97 and 0.99. A simple linear regression based approach for crowd estimation was applied to the data obtained within an environment in which the maximum occupancy as determined by the ground truth was approximately 3200 people. This led to median estimation errors between 191 and 286 individuals, depending on the data that was selected for training. The real-time RF-attenuation-based data we were able to provide to the organizers of the events at which we were permitted to install our system, was already considered to be particularly useful in a security context. However, the results we obtained have opened up a wide variety of future research prospects, of which the development of a training-less crowd estimation approach and the combination of our system with other data sources are considered to be one of the most important.

References

[1] M. W. Aziz, F. Naeem, M. H. Alizai, and K. B. Khan, "Automated solutions for crowd size estimation," *Social Science Computer Review*, vol. 36, no. 5, pp. 610–631, 2018.

[2] S. A. M. Saleh, S. A. Suandi, and H. Ibrahim, "Recent survey on crowd density estimation and counting for visual surveillance," *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 103–114, 2015.

[3] O. Kaltiokallio, R. Jäntti, and N. Patwari, "Arti: An adaptive radio tomographic imaging system," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7302–7316, 2017.

[4] B. Wei, W. Hu, M. Yang, and C. T. Chou, "Radio-based device-free activity recognition with radio frequency interference," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. ACM, 2015, pp. 154–165.

[5] S. Kianoush, S. Savazzi, F. Vicentini, V. Rampa, and M. Giussani, "Device-free rf human body fall detection and localization in industrial workplaces," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 351–362, 2016.

[6] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013, pp. 27–38.

[7] M. Youssef, M. Mah, and A. Agrawala, "Challenges: device-free passive localization for wireless environments," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. ACM, 2007, pp. 222–229.

[8] Y. Yuan, J. Zhao, C. Qiu, and W. Xi, "Estimating crowd density in an rf-based dynamic environment," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3837–3845, 2013.

[9] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only wifi power measurements," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 7, pp. 1381–1393, 2015.

[10] S. Di Domenico, M. De Sanctis, E. Cianca, P. Colucci, and G. Bianchi, "Lte-based passive device-free crowd density estimation," in *Communications (ICC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.

[11] S. Di Domenico, M. De Sanctis, E. Cianca, and G. Bianchi, "A trained-once crowd counting method using differential wifi channel state information," in *Proceedings of the 3rd International on Workshop on Physical Analytics*. ACM, 2016, pp. 37–42.

[12] S. Denis, R. Berkvens, B. Bellekens, and M. Weyn, "Large scale crowd density estimation using a sub-ghz wireless sensor network," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2018, pp. 849–855.

[13] S. Kianoush, M. Raja, S. Savazzi, and S. Sigg, "A cloud-iot platform for passive radio sensing: Challenges and application case studies," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3624–3636, 2018.

[14] *D7A Specification 1.1 Wireless Sensor and Actuator Network Protocol*, DASH7 Alliance, 2016, version 1.1.