EM-ID: Tag-less Identification of Electrical Devices via Electromagnetic Emissions

Chouchang Yang
Disney Research
Pittsburgh, PA 15213
email: jack.yang@disneyresearch.com

Alanson P. Sample
Disney Research
Pittsburgh, PA 15213
email: alanson.sample@disneyresearch.com

Abstract—Radio Frequency Identification technology has greatly improved asset management and inventory tracking. However, for many applications RFID tags are considered too expensive compared to the alternative of a printed bar code, which has hampered widespread adoption of RFID technology.

To overcome this price barrier, our work leverages the unique electromagnetic emissions generated by nearly all electronic and electromechanical devices as a means to individually identify them. This tag-less method of radio frequency identification leverages previous work showing that it is possible to classify objects by type (i.e. phone vs. TV vs. kitchen appliance, etc). A core question is whether or not the electromagnetic emissions from a given model of device, is sufficiently unique to robustly distinguish it from its peers.

We present a low cost method for extracting the EM-ID from a device along with a new classification and ranking algorithm that is capable of identifying minute differences in the EM signatures. Results show that devices as divers as electronic toys, cellphones and laptops can all be individually identified with an accuracy between 72% and 100% depending on device type.

While not all electronics are unique enough for individual identifying, we present a probability estimation model that accurately predicts the performance of identifying a given device out of a population of both similar and dissimilar devices. Ultimately, EM-ID provides a zero cost method of uniquely identifying, potentially billions of electronic devices using their unique electromagnetic emissions.

I. INTRODUCTION

The 5¢ RFID tag was famously claimed to be the tipping point for the RFID industry that would lead to wide spread adoption of the technology. While many of the innovations in cost reduction Sanjay Sarma outlined in 2001 [1] have come to pass, the 5¢ price point has remained elusive at any quantity. Presently, EPC Gen2 tags (ISO-18000-6c) can be readily purchased individually for ∼90¢ online and industry trade magazines report, that in large volumes, tags can be purchased in the 7¢ to 15¢ range [2]. While both academia and industry continue to make strides to reduce the cost of traditional UHF RFID tags, the reality is that due to inflationary forces the opportunity for a 5¢ tag may have already passed us by.

In an effort to further reduce cost and tag complexity, researchers are focusing on developing chip-less [3], [4] and antenna-less [5], [6] RFID tags. The general rule of thumb being that; a third of the cost of the tag is the integrated circuit (IC), the second third is the antenna inlay, and the final third represents the cost of bonding the two elements together.

Fig. 1. Depiction of an IT professional scanning the unique electromagnetic noise generated by the unmodified laptops in order to determine their unique IDs and recover their asset management information.

Therefore by eliminating one of these elements it should be possible to further reduce the cost of RFID tags.

However, all of these RF methods of uniquely identifying an object rely on adding some form of tag. For many applications, RFID is simply dismissed as being too expensive compared to the alternative of using optical identification in the form of printed barcodes or QR codes. While barcodes are often considered to be "free" they do take up valuable real estate on print media and/or require printed stickers that are manually applied to objects; both of which represent non-zero costs. This begs the question: *Can Radio Frequency Identification ever cost less then a barcode?*

This paper argues *yes it can*; and in fact there exists a subset of over a billion electronic devices that already have a unique radio frequency identity and are simply waiting to be read. We introduce a tag-less Radio Frequency Identification method that uses the unique electromagnetic signatures emitted by electronic devices as a means to identify individual objects, even of the same type and model.

Traditionally Electro-Magnetic Emissions (EME) have been simply thought of as incidental system noise that must be kept below a certain threshold to meet governmental regulations – the fact is that EME is highly structured and a direct mani-
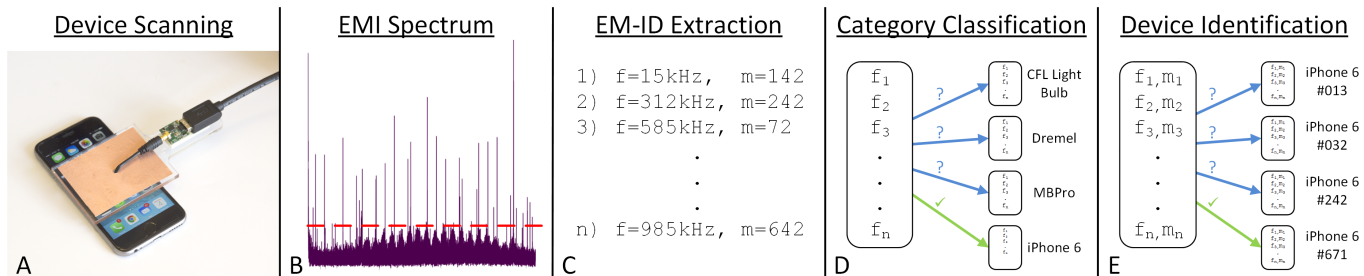
Fig. 2. Diagram depicting the process of uniquely identifying an electronic device (in this case a unknown iPhone 6) based on its low frequency electromagnetic emissions. Panel A shows a low cost EM-ID reader recording the EM signature shown in panel B. This signature is thresholded to remove noise and the device's EM-ID is extracted as depicted in panel C. In order to identify what type of device it is, a classification algorithm is done in panel D. Once the object type is known, it is compared to all other iPhone 6s in the database in order to recover its unique ID, as shown in panel E.

festation of the system circuits that generate it. Furthermore, variations in the manufacturing process at all levels, from the integrated circuits, to passive components, and board level layout all provide further EME differences between devices, even of the same model.

Figure 1 depicts an application scenario where an IT professional scans an unmodified electronic device (such as a laptop) to extract its unique EM-ID, which is then entered in a database. Since the EM-ID is persistent over time, the object can be scanned at a later date and its EM-ID is then compared to a database for identification. This unique identification method is free, in the sense that it already exists, cannot be removed, or be easily tampered with, but it does require that the device is powered on to generate the EME signal.

Previous work has shown that EME can be used to identify general classes of objects, for instance household appliances, computing devices, power tools, automobiles, etc. [7], [8]. However, using these methods it is not possible to uniquely identify objects of the same model nor is it possible to assign them individual IDs.

This work introduces a $10 electromagnetic emissions reader based on a software defined radio and a signal processing pipeline that is capable of robustly and repeatedly extracting the unique EM-IDs of individual devices. An overview of the system is presented in section II, along with a detailed description of our new signal processing and ranking algorithm used to identify individual objects in section IV. Since the EME signals emitted by a device is an emergent property and not simply a fixed digital number, section V presents a mathematical description of the probability distribution of the EME of a device and our ability to disambiguate similar devices in terms of the probability of successful identification. Lastly, experimental results are presented in section VI that show a wide range of devices of different complexity levels – from toys to cell phones and laptops – can all be uniquely and repeatedly read and identified.

II. SYSTEM OVERVIEW

This section provides a general overview of the process of scanning, classifying and identifying individual devices based on their electromagnetic signatures. Subsequent sections provide a deep dive into the algorithms needed for fine-grained object identification, and details on predicting the identification performance of a given set of devices.

A. EM-ID Reader Hardware

While lab equipment such as spectrum analyzers and high-speed oscilloscopes have traditionally been used to capture the electromagnetic signatures emitted by electronic and/or electromechanical devices, this work utilizes a software defined radio module based on the RTL-SDR, which works in conjunction with a smartphone or laptop to form a portable and low cost EM-ID scanner.

The RTL-SDR [9] is sold online for $10 (USD) and is based on the Realtek RTL2832 chip which performs I-Q demodulation and digitizes the IF signal with high speed 8-bit ADCs. The system streams raw data to a host computer via USB. In order to sample the low frequency EMI, the RF front end chip was removed and a WD2142 transformer is used to feed the raw EMI into the RTL2832 as previously described in [7].

The internal digital mixer and low-pass filter provides a selectable frequency window from 0 to 28.8 MHz at a maximum sampling rate of 3.2 MHz. Since most EM signals exist at the low frequencies of this range, the sampling rate is set to 1 MHz allowing the system to observe EM signals from $0 \sim 500$ kHz. Figure 2 panel A, shows a smartphone being scanned. An antenna consisting of an electrically short monopole is used to capture the EMI signals, which are fed to the modified RTL-SDR, and then processed by a host PC.

B. EM-ID Extraction

Once the EMI signal is digitized and sent to the host PC, it is converted to the frequency domain as shown in figure 2, panel B. In order to extract an EM-ID from the raw FFT data the low magnitude noise must be removed. This is done by setting a threshold, which is 1% higher then the difference between the peak and average signal magnitude: $Threshold = (peak - mean) \cdot 1\% + mean)$. Data points above the threshold are stored in an array of frequency and magnitude pairs that represents the EM-ID of the device, as shown in figure 2, panel C. While the number of frequency peaks is dependent on the

device type, typical EM-IDs have a length of 1,000 to 2,000 elements.

Once the EM-ID of a device has been recorded it can be stored in a central database along with other asset management information such as asset number, manufacturer, model, owner, etc. Since an EM-ID is based on random variations in the manufacturing of the electronic device it cannot be known *a priori*. Thus each device must be registered in a database, which is a typical process for most RFID applications.

C. Category Classification and Device Identification

As with all RFID systems, the ID number is simply a means of linking a physical object to a database of attributes, or to trigger an event/action. Since the EM-ID cannot store custom user information, the challenge is to link the EM-ID of a unknown scanned object to the correct entry stored in the database of EM-IDs. The process of quickly searching a database is more complicated for EME based identification since the EM-ID is an emergent statistical property of the system rather then a digitally stored ID number.

As will be described in detail in section IV the EM-ID is the result of a probability distribution. While it will be shown that the probability distribution from one device to the next is indeed unique and non-overlapping (and thus sufficient for identification), for the purpose of this system overview it is important to remember that from one read to the next there may be a small perturbation in a device's EM-ID. Therefore, to uniquely identify an individual device (that exists in the database) there is a two-stage ranking process starting with category classification and then device identification.

The goal of the category classification stage is to determine what type of device the unknown object is, thereby greatly reducing the search space for the identification stage. As described in Laput et. al. [7] category classification based on support vector machines (SVM) can be done robustly across a wide variety of devices. In this paper, we implemented a ranking system based on the cosine similarity function, which greatly reduces the computation complexity and eliminates the need for training. As shown in figure 2, panel D, the frequency components of the EM-ID of the unknown iPhone 6, is compared to examples of each of the subcategories. The results are ranked and the high matching categories are flag for further examination in the device identification stage.

To give some intuition into the effectiveness of this approach figure 3 shows the frequency distribution of five example devices (MacBook Pro, toy lightsaber, florescent lightbulb, LCD Screen, and an iPhone 6). Visually, it is easy to see that the spikes along the 500kHz frequency range are identifiable and unique to each object. Once this data is thresholded, the resulting frequency and magnitude pairs are recorded and robust similarity scoring can be done.

The final stage of the process is device identification which is depicted in figure 2, panel E. Here the goal is to disambiguate one device from others of the same type and model; for instance one iPhone 6 from a population of iPhone 6s. This is a much more challenging task, as the frequency distribution
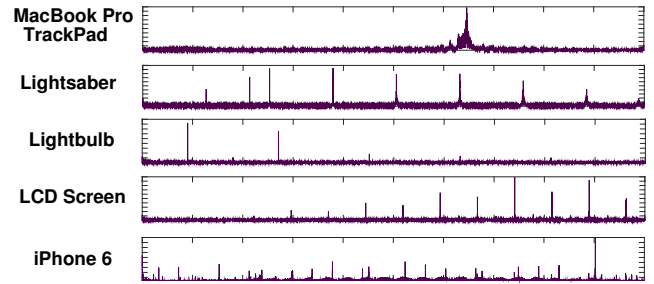

Fig. 3. Representative frequency domain plots from 0 to 500kHz of the EM signals from five different categories of devices. The distinctive patterns are a manifestation of the circuit topology and are unique to each device category.

will have a much higher likelihood of overlapping. As will be described in detail in section IV, both the frequency and magnitude components of the EM-ID will be used along with the cosine similarity function to uniquely determine the identity of the unknown objects.

III. MITIGATING EM-ID READER VARIATIONS

As with all radio receivers, different EM-ID readers have variations in signal sensitivity as well as different local oscillator precisions and offsets. Even when using the same EM-ID reader to rescan a given device, variations in reader antenna placement and distance can result in differences in the absolute magnitude of the EM signal. These variations can result in measurement uncertainty in both frequency and magnitude. The following section presents methods for mitigating these issues, thus allowing multiple EM-ID readers to be used interchangeably.

A. Frequency Uncertainty

Since the local oscillator (LO) on the EM-ID reader is used as a reference to measure the frequency components of the received electromagnetic emissions from the device under test, variations in the LO over time or from reader to reader, will result in mis-measured frequencies. For example, in this work, the EME signals are sampled at a rate of 1 MHz and an FFT window size of 2^{17} is used when converting from the time domain to the frequency domain. This results in a frequency resolution (or frequency bin width) of 7 Hz. However, the local oscillator used on the RTL-SDR has an accuracy of only 10 PPM (parts per million) and at 1MHz this results in an uncertainty of 10 Hz. Thus, it is possible that a signal source of 100 Hz will be measured by one reader at 110 Hz and another reader at 90 Hz, resulting in a mismatch between the respective frequency indexes in the recorded EM-ID. While changing system parameters such as FFT window size, sampling rate, and oscillator accuracy can improve the system, the underlying problem still exists. In typical radio systems this issue is overcome by having the transmitter send a pilot tone that the receiver's Phase Lock Loop can lock on to. However, no such signal exists for an EM-ID reader due to the emergent nature of EMI signals.

To mitigate the LO issue, a pseudo pilot tone approach is used where the frequency component with the maximum signal strength of the first EM-ID is used to align to the second EM-ID. Since the amount of frequency uncertainty is known, only peaks that are within +/-10 Hz are used for alignment, thus greatly reducing false matches for unrelated frequency spikes. For example, consider an EM signal which has multiple tones scattered over the frequency spectrum, two different EM-ID readers may measure the strongest frequency tone differently at 100 Hz and 110 Hz. The frequency compensation algorithm can shift all the 110Hz data by a -10 Hz offset. Therefore, after frequency offset compensation, two different measurements reach a consensus in terms of frequency measurements such that later their similarity can be evaluated.

B. Magnitude Uncertainty

The magnitude of recorded EM signals can very significantly based on the EM-ID reader's sensitivity and the distance and placement of the reader antenna to the device under test. To deal with this issue, the signals are normalized to a unit vector for similarity evaluation in later sections. Generally speaking, measurements can be taken robustly without the user worrying about aligning the antenna as long as it is touching or is tapped on the object of interest. For instance when scanning an iPhone, it simply needs to be placed on the reader. Alternatively when scanning a MacBook Pro track pad, the EM-ID reader antenna should touch the pad and not the LCD screen, as that would be a different measurement. This does make the reasonable assumption that there is not significant frequency dependent attenuation caused by user antenna placement or drifts in the analog front end of the reader.

IV. FINE-GRAINED CLASSIFIER FOR SIMILARITY MEASURES

Similarity measures between datasets have been studied in several fields [10]. These techniques include Euclidian distance, cosine similarity and relative entropy – these approaches quantify the similarity of two vectors in high dimensional data space [11], [12]. For example, cosine similarity computes the score between two vectors [13], [14]; the higher the score of cosine similarity, the more similar the datasets. In this section, a two stage cosine similarity algorithm is used to first classify the category of an unknown device and then determine its true identity by correctly determining which entry in the EM-ID database it belongs to.

A. Cosine Similarity

Cosine similarity gives a score based on the similarity of two vectors in higher dimensional space. Consider a time series EM signal $\mathbf{x} = [x_1, x_2, x_3, ..., x_n]$ sensed by an EM-ID reader. The frequency transform of this EM signal is $\mathbf{X} = \mathscr{F}\{\mathbf{x}\}$ and represents the EM signal's frequency distribution. The vector $\mathbf{X} = [X_1, X_2, X_3, ..., X_N]$ represents the EM signal data in the frequency domain where each element X_i ($0 \leq i \leq N$) refers to the signal strength at the i_{th} frequency bin. Similarly,

another unknown EM signal after frequency transformation is $\mathbf{Y}$. By viewing each EM signal $\mathbf{X}$ and $\mathbf{Y}$ as n-dimensional vectors, the cosine similarity yields a score calculated by:

$$\text{C.S.}(\mathbf{X}, \mathbf{Y}) := \frac{\mathbf{X} \cdot \mathbf{Y}}{\|\mathbf{X}\| \, \|\mathbf{Y}\|} = \frac{\sum_{i=1}^{N} X_i Y_i}{\sqrt{\sum_{i=1}^{N} X_i^2} \sqrt{\sum_{i=1}^{N} Y_i^2}} \quad (1)$$

A high value of cosine similarity corresponds to two similar datasets. Since each element X_i and Y_i for $0 \leq i \leq N$ represents absolute magnitude, the values X_i and Y_i are always positive – this results in a cosine similarity range between 0 to 1. A similarity score of 0 implies two vectors are orthogonal to each other, while a score of 1 indicates the two vectors are identical. Given two known EM signals $\mathbf{X}$ and $\mathbf{Y}$ obtained from an electronic object "E" and "G" respectively, and an unknown EM signal $\mathbf{X}'$, using cosine similarity we can identify which object the unknown vectors corresponds to. If $\text{C.S.}(\mathbf{X}, \mathbf{X}') > \text{C.S.}(\mathbf{Y}, \mathbf{X}')$, then the cosine similarity indicates the unknown EM signal X' is from the same source as X. Hence, the result identifies the unknown signal $\mathbf{X}'$ as the electronic object "E". Similarly, if $\text{C.S.}(\mathbf{X}, \mathbf{X}') < \text{C.S.}(\mathbf{Y}, \mathbf{X}')$, the unknown EM signal $\mathbf{X}'$ is identified as the electronic object "G".

1) Gain difference vs. Similarity:
Equation (1) implies that all gain differences can be mitigated because each vector $\mathbf{X}$ and $\mathbf{Y}$ are normalized by their total energy $\|\mathbf{X}\|$ and $\|\mathbf{Y}\|$ respectively. To illustrate this concept, consider a time series EM signal $\mathbf{x}'$ that has α times more gain than $\mathbf{x}$ ($\mathbf{x}' = \alpha\mathbf{x}$). By applying the fact that the same gain exists in the frequency domain as $\mathbf{X}' = \alpha\mathbf{X}$, the cosine similarity can be obtained by

$$\text{C.S.}(\mathbf{X}', \mathbf{Y}) := \frac{\mathbf{X}' \cdot \mathbf{Y}}{\|\mathbf{X}'\| \, \|\mathbf{Y}\|} = \frac{\alpha\mathbf{X} \cdot \mathbf{Y}}{\alpha \|\mathbf{X}\| \, \|\mathbf{Y}\|} = \text{C.S.}(\mathbf{X}, \mathbf{Y})$$
$$(2)$$

Equation (2) indicates that the signal strength, in terms of gain coefficient, will be normalized through the cosine similarity calculation. Thus, once the EM signals for given devices are established in a database, the cosine similarity can evaluate each new unknown EM signals even when measured in different positions and orientations, or by different EM-ID readers.

B. Category Classification

When EM signals are from different device categories, most of their frequency bins do not overlap. This can be seen in figure 3, where the Macbook Pro trackpad, lightsaber, florescent lightbulb, LCD screen, and iPhone 6 have completely different frequency distributions. Since their internal circuit structures are completely different, different categories have no or few common frequency bins. When two vectors have no common frequency bins, the cosine similarity returns to zero. For example, two EM signals from different categories with frequency distribution $\mathbf{X} = [X_1, X_2, X_3, 0, 0, ..., 0]$ and
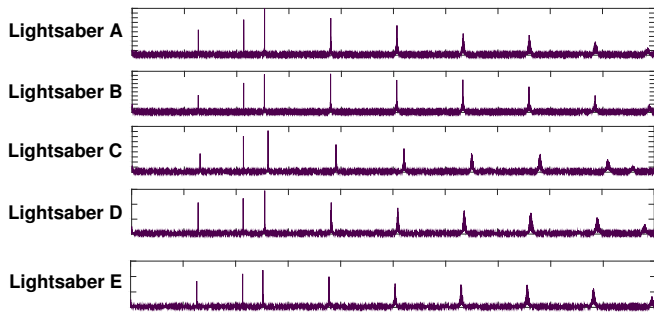
Fig. 4. Frequency domain plots from 0 to 500kHz of the EM signals for five identical toy lightsabers. Since the same circuit is used in each device the frequency distributions are similar to each other, although close inspection shows that indeed there are small differences.

$\mathbf{Y} = [0, 0, 0, Y_1, Y_2, Y_3, 0, 0..., 0]$ have zero scores. If EM signal vectors are from different objects but are within the same category, their frequency distributions usually have a higher fraction of common frequency bins because of their similar circuitry. The lighsaber example shown in figure 4 yields a very similar frequency distribution and a higher fraction of overlapping frequencies. Consider two EM signal vectors from the same categories as $\mathbf{X} = [X_1, X_2, X_3, 0, 0, ..., 0]$ and $\mathbf{Y} = [0, Y_1, Y_2, Y_3, 0, 0, ..., 0]$, their cosine similarity is $\frac{X_2 Y_1 + X_3 Y_2}{\|\mathbf{X}\|\|\mathbf{Y}\|}$ and returns a non-zero score. As a result, different objects within the same category can have a higher cosine similarity than objects from different categories. By leveraging this property, we can first identify the class of device by using a known vector for each category. For example, comparing an EM signal of an unknown device to representative EM-ID vectors from each category of devices in the database will result in a set of similarity scores. The highest score will correspond to the category of the unknown device which can then be used to narrow the search space when determining the identity of the object among objects of the same category and model in the database.

C. Fine-grained Identification

Once the class of a device is determined, it is then compared to all known objects in the database of the same class using the cosine similarity function to determine its identity. The highest score is used to return the object's true ID. For example, an unknown object is first compared to one representative of each device class in the EM-ID database. If the highest scoring EM-ID vector is the lightsaber class then the unknown devices is compared to all lightsabers in the EM-ID database. The system will then return the ID of the lightsaber with the highest probability match based on the cosine similarity function.

V. PREDICTING IDENTIFICATION SUCCESS RATE

Since the electromagnetic emissions from electronic devices are an emergent property of the system and not specifically designed to be unique, there is the possibility that the EM emission spectrums of closely related devices will overlap causing identification errors. To investigate this issue we employ the Euclidean distance function [15] to analyze a large population of EM signatures for a set of devices, and then by modeling them as a Gaussian distribution we are able to calculate the probability of successfully identifying a given device.

A. Performance Analysis via Euclidean distance

When comparing two vectors the Euclidean distance function computes the same information as the cosine similarity function but its output is a linear vector rather then angle which makes it more applicable for plotting and manipulating probability distributions. Given two EM-IDs represented by an n dimensional vector of frequency and magnitude pairs $(\mathbf{X}, \mathbf{Y})$. The Euclidean distance can be calculated by first taking the unit vector of each $\mathbf{u_x} = \frac{\mathbf{X}}{\|\mathbf{X}\|}$, $\mathbf{u_y} = \frac{\mathbf{Y}}{\|\mathbf{Y}\|}$ and then calculating the distance between them $d(\mathbf{u_x}, \mathbf{u_y}) = \|\mathbf{u_y} - \mathbf{u_x}\|$. Smaller values of $d(\mathbf{u_x}, \mathbf{u_y})$ represent EM-ID vectors that are closely aligned, while larger distance numbers indicate vectors that are dissimilar.

We now return to the previous example of five identical toy lightsabers as shown in figure 4 which consists of five known EM signals $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{E}$. With the goal of showing how robustly lightsaber $\mathbf{A}$ can be identified out of the population of five devices; 300 scans of test data are taken of lightsaber $\mathbf{A}$ and denoted by $\hat{\mathbf{A}}_i$ where $i = 1, 2, ..., k$ represents each trial. Computing the Euclidean distance of all 300 trials of $\hat{\mathbf{A}}_i$ across the five known elements in the data base ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{E}$) results in 1,500 similarity measurements. These measurements are then normalized and plotted as the probability histogram depicted in figure 5. For instance the red block of histogram data shows all 300 distance measurements between lightsaber $\hat{\mathbf{A}}_i$ and $\mathbf{A}$ (i.e. measurements against itself). As expected the distance between the test data $\hat{\mathbf{A}}_i$ and $\mathbf{A}$ is smaller then the distances reported for the other lightsaber, $\mathbf{B}$ maroon, $\mathbf{C}$ green, $\mathbf{D}$ blue and $\mathbf{E}$ khaki. Furthermore since the red histogram for $\mathbf{A}$ does not overlap with any of the other lightsabers this shows that it is sufficiently unique, compared to its peers, such that it can be robustly identified without errors.

While the toy lightsaber is an example of a class of devices where it is easy to identify individual instances of the given device this is not always the case. As can be seen in figure 6, which shows a histogram of the measured Euclidean distance of five identical Apple Macbook Pros. As in the previous example 300 test scans of Macbook Pro B's trackpad ($\hat{\mathbf{B}}_i$) was scanned and the Euclidean distance between it and each of the five Macbook Pros in the database ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{E}$) are calculated and plotted resulting in 1,500 data points. Ideally, MacBook Pro $\mathbf{B}$ (shown as a green histogram) should have the smallest distance when compared to others four MacBook Pros. However, some of MacBook Pro's $\mathbf{E}$ histogram (shown in khaki) overlaps with the MacBook Pro $\mathbf{B}$ which results in an increased probability of identification errors. These two examples show that the higher the degree of overlap of the histogram (i.e. the higher the similarity between devices) the higher the probability of classification and identification errors. Likewise if the histograms do not overlap and have wide margins the probability of errors is quite low.
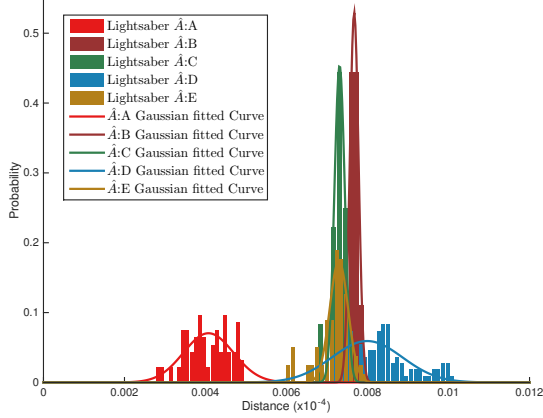
$$\mathrm{P_S} = \mathrm{Prob.}[\, d(\mathbf{u_x}, \mathbf{u_{x'_i}}) < d(\mathbf{u_y}, \mathbf{u_{x'_i}})\,] \qquad (3)$$

Equation 3 presents the probability of success by finding all distances from the same device that are smaller than those from different devices. Therefore, once μ_S, σ_S, μ_D and σ_D are obtained, the success rate $\mathrm{P_S}$ can be calculated through the Gaussian distribution model for each electronic device as:

$$\mathrm{P_S} = \int_{-\infty}^{\infty} \int_{x}^{\infty} \frac{1}{\sigma_S \sigma_D 2\pi} e^{-\frac{(x-\mu_S)^2}{2\sigma_S^2}} e^{-\frac{(y-\mu_D)^2}{2\sigma_D^2}} \, dxdy \qquad (4)$$

$$\approx \int_{\mu_S-3\sigma_S}^{\mu_S+3\sigma_S} \int_{x}^{\mu_D+3\sigma_D} \frac{1}{\sigma_S \sigma_D 2\pi} e^{-\frac{(x-\mu_S)^2}{2\sigma_S^2} - \frac{(y-\mu_D)^2}{2\sigma_D^2}} \, dxdy$$

Figure 7 illustrates the probability of accurately classifying an unknown device when the database contains overlapping Gaussian probability distributions. The blue curve represents the $d(\mathbf{u_{x'_i}}, \mathbf{u_x})$ distribution while the brown curve represents the $d(\mathbf{u_{x'_i}}, \mathbf{u_y})$. The area shaded in blue indicates a True Positive identification, where an unknown device is correctly classified as matching the test vector stored in the database. The area shaded in brown indicates a True Negative identification, where an unknown device is correctly classified as <u>not</u> matching the test vector stored in the database. The region of space the falls under both the blue and brown curves that is shaded in red is a concatenation of the False Negative and False Positive identification, and represents an error. Thus the more the two Gaussian distributions overlap, and the larger the red region, the higher the probability of miss classifying the unknown device. Therefore to calculate the probability of successfully classifying an unknown device, one must calculate the area of the blue and brown shaded regions using equation 4.

Since 99.73% of the Gaussian distribution area is within 3 standard deviations of distance, the integration area can be reduced to obtain the approximate probability of success. Thus, this leads to one important result: a 100% success rate, or an error-free identification, must have the following property:

$$\mathbf{ErrorFree} : (\mu_D - \mu_S) > 3 \times (\sigma_D + \sigma_S) \qquad (5)$$

To achieve error free identification as defined in 5, the relative distance of the mean between the blue and brown curves must be more than 3 total sigma ($\sigma_D + \sigma_S$). This property can be used to predict why the lightsaber is error free while the Macbook Pro trackpad's EM signals have some identification failures. As will be shown in the results section, the equality $(\mu_D - \mu_S) > (3 \times (\sigma_D + \sigma_S))$ is true for the lightsaber category, but not the Macbook Pro category.

VI. EXPERIMENTAL RESULTS AND PERFORMANCE

In order to evaluate the effectiveness of the proposed EM based identification system both category and individual device identification has been tested. Additionally the success rate prediction algorithm is also used to estimate performance which is then compared to measured results.

Fig. 5. A histogram of Euclidean distances between an unknown test lightsaber $\hat{\mathbf{A}}_i$ and the five known lightsabers **A**, **B**, **C**, **D** and **E** who's EM-IDs are stored in a database. In these example lightsaber **A** is correctly classified as matching $\hat{\mathbf{A}}_i$ since it has the shorted distance and its probability distribution does not overlap with any other lightsabers.
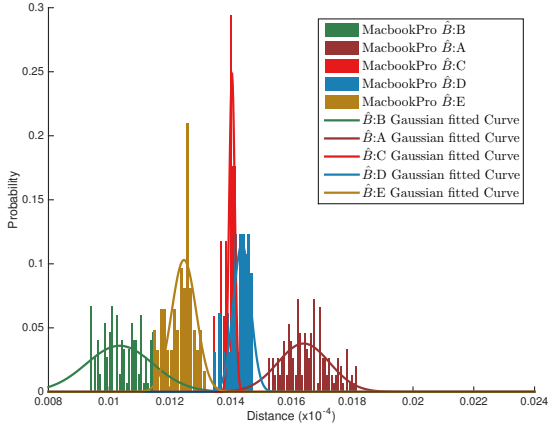

Fig. 6. The histogram distances distributions of EM signal are constructed through Macbook Pro trackpad. The distances of the pair of EM signals collected from Macbook Pro **B** has some overlapping areas with the distance between the pair of Macbook Pro $\hat{\mathbf{B}}_i$ and **E** such that the identification failure happens. The error performance can be predicted through Gaussian distribution models.

B. Gaussian Distribution Prediction

This section describes a method for quantifying the probability of successfully identifying a device based on the distribution of Euclidean distances. By modeling the histogram of distance data as a Gaussian distribution, as shown in figure 5 and 6, the statistical information of all Euclidean distances can be used to predict the systems ability to correctly identify objects based on their EM signal from a database of EM-IDs. Consider μ_S and σ_S, which are the mean and standard deviation of all distances from the same electronic device $d(\mathbf{u_x}, \mathbf{u_{x'_i}})$, while μ_D and σ_D are the mean and standard deviation for all distances from different devices but within the same category $d(\mathbf{u_{x'_i}}, \mathbf{u_y})$. The probability of success can be computed as:
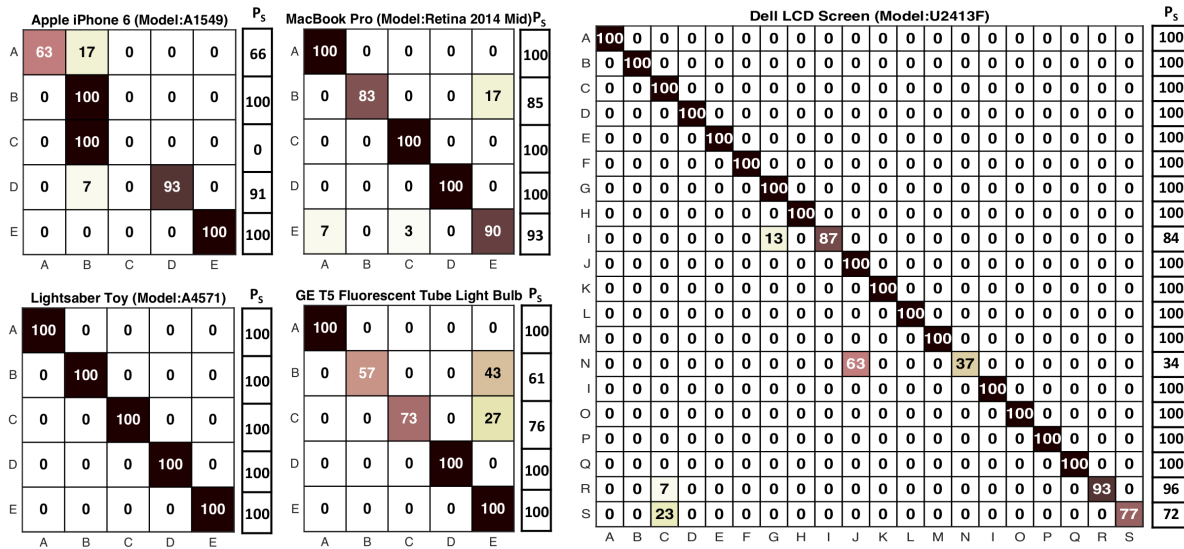
Fig. 9. Confusion matrices showing the likely hood that a particular device can be positively identified out of a population of devices of the same model. Note that the column denoted P_S next to each matrix shows the predicted success rate of identifying each individual device.

Apple iPhone 6 (Model:A1549)

	A	B	C	D	E	P_S
A	63	17	0	0	0	66
B	0	100	0	0	0	100
C	0	100	0	0	0	0
D	0	7	0	93	0	91
E	0	0	0	0	100	100

MacBook Pro (Model:Retina 2014 Mid)

	A	B	C	D	E	P_S
A	100	0	0	0	0	100
B	0	83	0	0	17	85
C	0	0	100	0	0	100
D	0	0	0	100	0	100
E	7	0	3	0	90	93

Lightsaber Toy (Model:A4571)

	A	B	C	D	E	P_S
A	100	0	0	0	0	100
B	0	100	0	0	0	100
C	0	0	100	0	0	100
D	0	0	0	100	0	100
E	0	0	0	0	100	100

GE T5 Fluorescent Tube Light Bulb

	A	B	C	D	E	P_S
A	100	0	0	0	0	100
B	0	57	0	0	43	61
C	0	0	73	0	27	76
D	0	0	0	100	0	100
E	0	0	0	0	100	100

Dell LCD Screen (Model:U2413F)

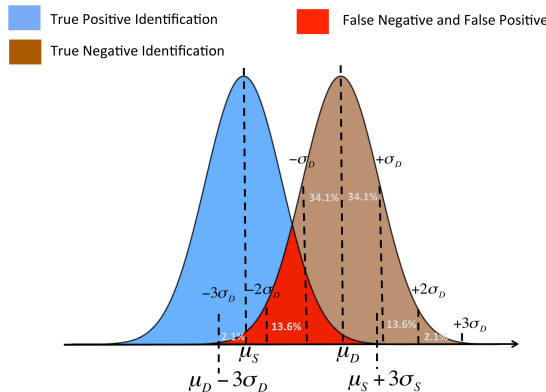|	A	B	C	D	E	F	G	H	I	J	K	L	M	N	I	O	P	Q	R	S	P_S
A	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
B	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
C	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
D	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
E	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
F	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
G	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	100
H	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	100
I	0	0	0	0	0	0	13	0	87	0	0	0	0	0	0	0	0	0	0	0	84
J	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	100
K	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	100
L	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	100
M	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	100
N	0	0	0	0	0	0	0	0	63	0	0	0	0	37	0	0	0	0	0	0	34
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	100
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	100
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	100
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	100
R	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	93	0	96
S	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	77	72

Fig. 7. The successful rate of identification can be estimated through the integral of true positive and true negative.

For testing, five different categories of electronic devices were chosen ranging from simple toys to laptop computers. For each device type identical versions of the same model were used for testing the system's ability to disambiguate similar
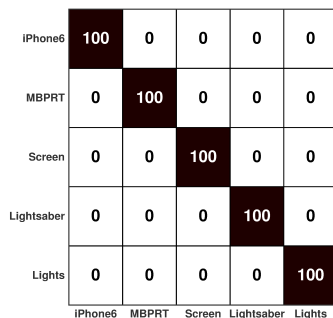

Fig. 8. Confusion matrix for category classification, showing that the EM-ID algorithm can reliably determine device type based solely on its electromagnetic emissions.

electronic devices. A total of 40 devices were tested including: 5x General Electric fluorescent tube light bulb (Model: GE-F54W), 5x Hasbro lightsaber toy (Model: A4571), 20x Dell LCD 24 inch screen (Model: U2413F), 5x Apple iPhone 6 (Model: A1549), and 5x Macbook Pro Retina Mid-2014. For both the iPhone 6 and Dell LCD screens the same image was displayed on each of the five respective units during EM measurements. For the five Macbook Pros the trackpad was scanned with the EM-ID reader, although other parts of the laptop such as the screen and keyboards could also be used as a secondary measurement for redundancy.

To generate the EM-ID database each of the 40 devices was scanned once and their respective EM-IDs where stored. To show that the EM emissions from these devices are stable and persistent overtime the EM-ID database was collected on October 27, 2015 and the testing was conducted five months later on March 14, 2016. In order to show the reliability of the EM-ID system to both determine a device's category and its individual identity, each of the 40 devices was scanned 30 times and compared to EM-IDs in the database, resulting in 1,200 total trials.

The results from category testing are shown in figure 8. In order to reduce computational complexity only one EM-ID from each of the five categories was used for comparison. For example when testing an unknown device, only one device from each category was needed for comparison. Results show that the cosine similarity function produced robust category classification accuracy across the 1,200 trials with a total accuracy of 100%.

Once an unknown device's category has been determined the second, and more challenging, task is to determine which of the five devices of the same model it corresponds to. Again, the cosine similarity function is used to determine the degree to which the EM-ID candidate vectors are aligned and the

one with the highest score is chosen as a match. Results are shown in the five confusion matrices in figure 9. Five identical models of the lightsaber, MacBook Pro, florescent lightbulb, and iPhone 6 were each tested for 30 trials. For the Hasbro Lightsaber the identification algorithm perfectly identified the individual device 30 times (i.e. Lightsaber A was correctly identified as Lightsaber A). The five MacBook Pros had an average identification accuracy of 94.6% and the GE florescent lightbulb had an average accuracy of 86%. Fortunately, 20 units of the Dell LCD screen was available for testing and each unit was scanned 30 times. The results shows and average identification accuracy of 94.7%.

The results for the iPhone 6 are less reliable with an average accuracy of 71.2%. This is primarily due to iPhone "C" being completely miss-categorized as iPhone "B". As described in the text since the electromagnetic emissions generated by a device are an emergent property it is not possible to ensure that EM-IDs are always unique and never collide. Fortunately the algorithm for *Predicting the Identification Success Rate* (presented in section V was able to accurately predict these failures. This is shown in the columns labeled P_S next to the confusion matrices shown in figure 9, which accurately predicts the success rate. This is an important result since the success rate prediction P_S was made based only on the original EM-ID database created in October. Thus when a user enters a device into the database the success rate prediction algorithm can be run and can alert the user if the new device is unique enough to be read or if an alternate strategy is needed.

Error free identification performance can be predicted by using the Euclidian distance information in Table I, which list each categories average distance and standard deviation. For instance, for the lightsaber μ_D - μ_S = 2.69×10^{-3} while $3 \times (\sigma_D + \sigma_S) = 1.098 \times 10^{-3}$. Thus by using the criteria for error free identification performance defined in equation (5), the lightsabers satisfies the equality μ_D - $\mu_S > 3 \times (\sigma_D + \sigma_S)$ such that it has error free identification performance.

TABLE I
AVERAGE EUCLIDEAN DISTANCE FOR EACH CATEGORY: (UNIT:10^{-3})

Category	iPhone6	MBPR	Screen	Lightsaber	Lights
Avg:μ_S	3.341	8.521	5.72	3.703	32.173
σ_S	0.33	0.2356	0.483	0.253	1.867
Avg:μ_D	3.452	10.571	8.124	6.393	33.406
σ_D	0.324	0.3325	0.351	0.113	1.763

VII. CONCLUSION

This paper proposes a method for individually identifying electronic devices, without the need to add bar codes are RFID tags, simply by measuring the electromagnetic noise they generate when powered on. These electromagnetic emissions are captured using a low cost (∼$10) hand held reader and can reliably extract the EM-ID of the device under test.

This work presents a new computationally lightweight algorithm that can determine the similarity in the EM-IDs of devices and shows an increase in category classification accuracy over previous work. More importantly for the first time, this work tackles the more challenging task of individually identifying an unknown device out of a population of the same model. Testing was done on electronic devices as simple as a light saber toy and as complex as a smart phone with identification accuracy ranging from 100% to 71.2% respectively.

While not all classes of electronic devices are guaranteed to be individually identifiable the success rate prediction algorithm presented here can accurately predict the likelihood that a device can be identified out of a population of the same model. Ultimately the signal processing and mathematical frame work established here, lays the foundation for utilizing electromagnetic emissions for identification purposes.

REFERENCES

[1] S. Sarma, "Towards the 5 tag," tech. rep., Auto-ID Center, 77 Massachusetts Avenue, Bldg 3-449, Cambridge, MA 02139-4307, USA, nov 2001.

[2] R. J. LLC, "How much does an rfid tag cost today?," sep 2015.

[3] S. Preradovic, I. Balbin, N. Karmakar, and G. Swiegers, "Multiresonator-based chipless rfid system for low-cost item tracking," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 57, pp. 1411–1419, May 2009.

[4] V. Plessky and L. Reindl, "Review on saw rfid tags," *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, vol. 57, pp. 654–668, March 2010.

[5] P. Pursula, T. Vaha-Heikkila, A. Muller, D. Neculoiu, G. Konstantinidis, A. Oja, and J. Tuovinen, "Millimeter-wave identification, a new short-range radio system for low-power high data-rate applications," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 56, pp. 2221–2228, Oct 2008.

[6] X. Chen, W. G. Yeoh, Y. B. Choi, H. Li, and R. Singh, "A 2.45-ghz near-field rfid system with passive on-chip antenna tags," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 56, pp. 1397–1404, June 2008.

[7] G. Laput, C. Yang, R. Xiao, A. Sample, and C. Harrison, "Em-sense: Touch recognition of uninstrumented, electrical and electromechanical objects," in *Proceedings of the 28th Annual ACM Symposium on User Interface Software Technology*, UIST '15, (New York, NY, USA), pp. 157–166, ACM, 2015.

[8] N. Zhao, G. Dublon, N. Gillian, A. Dementyev, and J. Paradiso, "Emi spy: Harnessing electromagnetic interference for low-cost, rapid prototyping of proxemic interaction," in *Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on*, pp. 1–6, June 2015.

[9] S. Cass, "A 40 dollar software-defined radio," *Spectrum, IEEE*, vol. 50, no. 7, pp. 22–23, 2013.

[10] G. P. Penney, J. Weese, J. Little, P. Desmedt, D. L. Hill, D. J. Hawkes, *et al.*, "A comparison of similarity measures for use in 2-d-3-d medical image registration," *Medical Imaging, IEEE Transactions on*, vol. 17, no. 4, pp. 586–595, 1998.

[11] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity," in *AAAI*, vol. 6, pp. 775–780, 2006.

[12] A. Huang, "Similarity measures for text document clustering," in *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*, pp. 49–56, 2008.

[13] R. J. Bayardo, Y. Ma, and R. Srikant, "Scaling up all pairs similarity search," in *Proceedings of the 16th International Conference on World Wide Web*, WWW 2007, (New York, NY, USA), pp. 131–140, ACM, 2007.

[14] S. Tata and J. M. Patel, "Estimating the selectivity of tf-idf based cosine similarity predicates," *SIGMOD Rec.*, vol. 36, pp. 7–12, June 2007.

[15] D. E. Holcomb, W. P. Burleson, and K. Fu, "Power-up sram state as an identifying fingerprint and source of true random numbers," *Computers, IEEE Transactions on*, vol. 58, no. 9, pp. 1198–1210, 2009.