Energy Aware Information Retrieval from Mobile Phones

ABSTRACT

On mobile phones, Information Retrieval (IR) applications that search over remote collections consume significant energy as a result of network activity. Typical IR applications mainly focus on retrieving and presenting relevant documents to users in an effective manner and do not take into account the cost of network data transfer. Despite the improvements in network access, energy consumption of network activity can seriously reduce battery lifetime constraining the use of IR applications. Therefore, on mobile phones in addition to being effective an IR application must also be energy efficient.

In this paper, we study the energy consumption of a typical IR application on a mobile phone for the task of retrieving documents residing on a remote server. Our measurements indicate that the network activity of typical IR process consumes a substantial amount of energy and can reduce battery life (up to 50%) when compared to a completely local IR application. To explore opportunities for saving energy, we first build a model for the energy consumption of network activity induced by the IR application. Motivated by the characteristics of this energy model, we explore two opportunities for reducing energy consumption of a typical IR application: 1] Reducing the overall amount of data transferred using relevance feedback and 2] Reducing the number of data transfers for each query by bundling more data at each transfer opportunity. The modifications we propose provide substantial savings in energy (up to 25%) with additional benefits and trade-offs incurred in dead time and latency. Actual implementations of the proposed modifications confirm our model based findings.

1. INTRODUCTION

Smart-phones and PDA's with their enhanced touch-screen displays and access to high-bandwidth data plans are becoming powerful mobile computing devices. Recent query log studies of top search engines indicate that the number of mobile users, and the number and length of search sessions from mobile devices are increasing [4, 8]. However, energy consumption of network activity remains a key challenge to the use of searching for remote content from a mobile phone.

Energy consumption of network activity on mobile phones has been studied for general data transfers, service discovery, location sensing and transmission of sensor information [11],[5]. Previous research on information retrieval from mobile phones has focused on interface issues and handling the display and low computing power constraints on mobile phones by adapting the search results and the layout of web pages [10],[12],[9],[6] and location and context aware search [3],[7],[1]. To the best of our knowledge there has not been any previous research that considered the *energy consumption characteristics* of a standard information retrieval process on mobile phones.

Consider the standard information retrieval process: The user issues a keyword query to which the retrieval engine responds with a ranked list of snippets, typically with hyperlinks to documents. The user peruses the snippets and then downloads one or more documents. The user requests additional result pages and repeats the process until the information need is satisfied or another query is entered. This process induces a significant amount of network activity in terms of the amount of data transferred and the number of transfers depending on the number of documents perused, the effectiveness of the retrieval system and the number of results displayed per page. However, we do not know how much energy is consumed per query under this process, how it affects battery life, and how improvements in effectiveness can impact energy.

Indeed, standard information retrieval algorithms and applications were not designed for energy constrained devices such as mobile phones. One of the main goals of the standard information retrieval process is to present relevant information to the user as quickly as possible without overwhelming the user. Consequently, the number of results per page and the amount of information presented in a snippet are often determined based on display constraints, UI factors and efficiency considerations such as latency of responses. However, these choices can directly affect the network activity in terms of the number of transfers and the amount of data that gets transferred and may adversely impact energy consumption. Modeling the network activity of a typical IR application and understanding the energy consumption of network activity on mobile phones can help to identify opportunities for reducing energy consumption.

In this paper, we study the energy consumption characteristics of standard information retrieval applications over a cellular data network (GSM/EDGE). Using data transfer experiments we build a simple model that can predict energy consumed for data transfers of a given size. Based on the characteristics of the model we identify two simple opportunities for reducing energy consumption of search. First, we can reduce energy consumption by reducing the overall data transferred by improving retrieval effectiveness. Second, we find that making fewer data transfers using bigger blocks of data is more energy efficient than making more data transfers using smaller blocks of data. Accordingly, we propose two modifications to the interaction model to include more data – more snippets or documents – at each transfer opportunity to reduce the overall energy consumption per query. Our model based evaluation demonstrates that the modifications we propose can result in significant energy savings (up to 25%) over the standard interaction model. Further analysis show that the modifications also result in additional benefits in terms of latency and dead time. Finally, we validate our findings by using actual runs of a search application from a mobile phone. We conclude the energy consumption of IR can be greatly improved by adapting IR to suit the energy model of network activity.

2. MOTIVATION

Network activity on mobile phones consumes energy. The rate and amount of energy consumption is dependent on a number of factors such as the underlying network interface - cellular versus Wi-Fi, the signal strength, network and channel congestion. Previous studies have shown that on GSM – a type of cellular data network, power consumption is an order of magnitude higher than Wi-Fi and increases much faster with increase in data sizes. Our initial energy measurements on a mobile phone (Figure 1) are consistent with these results. While Wi-Fi is very energy efficient, it is often not available, especially when mobile. Therefore, cellular data networks such as GSM/Edge are often used for performing network activities from mobile phones. So, in this work we focus on the energy consumption of GSM/Edge networks.

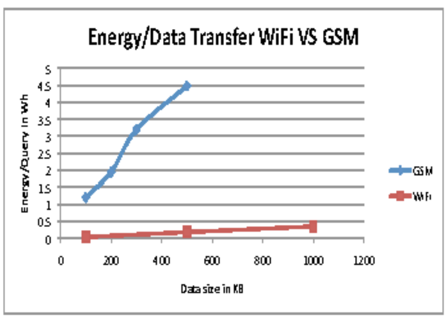

Figure 1: Energy consumption for data transfers: Wi-Fi versus GSM

To demonstrate the energy consumption and reduction in battery life due to network activity of an IR application, we first ran local retrieval experiments on the phone. We simulate a user who repeatedly issues queries, peruses snippets, selects relevant documents, reads them and moves on to the next query. When we switch to searching over a remote collection, the queries, snippets and documents are transferred over the cellular data network and incur additional energy thereby reducing battery life. Table 1 displays the energy consumption of a typical IR application using the queries and documents for the TREC Robust 04 collection(described later). Comparing the local versus network based search we see that there is almost a 50% reduction in battery life. Therefore, we focus on the network-based search over GSM setting where energy consumption is a critical issue: We build an energy model for information retrieval and explore avenues for reducing energy consumption.

3. AN ENERGY MODEL FOR IR

Our main goal is to model the average energy consumed in issuing a query and retrieving documents over the network. Under this model reductions in energy will correspond to an increase in the battery life when running the same number of queries. Alternatively, we can view the impact of energy savings in terms of the number of additional queries that can be run under an improved retrieval method. For example, a 10% decrease in energy consumed by method A over method B will mean that using method A we expect to run 10% more queries before the battery drains completely.

Method	Battery Life	Total Queries
Local	13.5 hours	800
Network	6.48 hours	368

Table 1: Energy consumption of a standard Information Retrieval Application: Searching for 5 Relevant Documents for each query in Robust 04 collection

The typical information retrieval process begins with a query issued by a user. First, the user is presented with a results page that consists of an ordered list of T snippets that serve to highlight the relevant portions of the corresponding documents. Then, the user peruses the snippets, guesses the relevance of the documents based on the snippets and selects a document to read. Iteratively, the user selects other documents from the presented result page to read. If necessary, the user will then proceed to request additional result pages and repeat the process until the information need is satisfied or another query is entered. For consistency of analysis, we will assume the following: The user is interested in reading a fixed number of relevant documents, r, for all queries and only selects relevant documents for downloads. The network activity induced for each query is a function of the number of result pages requested and the number of documents downloaded by the user. In fact, if the rth relevant document was retrieved at rank j, then the user must have downloaded at least $m = \lceil \frac{j}{10} \rceil$ result pages each of length p with 10 snippets, and r documents.

The energy consumed per query, $E(q)$, is simply the sum of energy consumed by all data transfers, $D(q)$ pertaining to the query:

$$E(q) = \sum_{d \in D(q)} E(d) \text{ Wh} \quad (1)$$

$$= m * E(p) + \sum_{i=1}^{r} E(|d_i|) \quad (2)$$

Using a reliable model that predicts energy consumption of data transfers, we are able to estimate the energy consumption of different information retrieval systems without having to run time consuming retrieval experiments on the mobile phone.

3.1 Measuring Data Transfer Energy

To model the energy consumption of network activity induced by an IR application, we first need to model the energy consumption of individual data transfers. To this end, we obtain energy measurements for data transfers of a wide range of data sizes. Ideally, we would like to measure remaining battery life before and after each data transfer to determine the energy consumed. However, it is hard to predict the remaining battery life accurately. For instance, the device used in this study provides remaining battery life measurement in a coarse seven point scale and is also not very reliable due to the non-linearity of battery drain. Instead, we use the following methodology: From the fully charged state, we repeat data transfers of a given size, waiting for 10 seconds after each transfer, until the battery drains completely. Based on the measured battery life and average transfer time, we estimate the average energy per data transfer as shown below.

The battery capacities are specified in terms of number of hours the battery can retain its electric potential while a constant current is drawn through it. For instance, a Nokia N95 phone has a specified battery capacity of 950 mAh, i.e., the battery will last for 950 hours if a steady current of 1milli-amperes is drawn from it. If we assume unit time to be 1 second, then the total energy that the battery can supply at a voltage of V_c with a draw of 1mA is $(950) * 0.001 * V_c$. Then, a reduced battery life l for a particular data transfer experiment implies that a higher current I is drawn through the battery during each time instant of the data transfer (assuming at all other times no current is drawn from the battery). Since the cumulative energy that the battery provides is the same, we must have the current drawn through the battery during each transfer time instance as $I = \frac{950*0.001*V_c}{l*V_c} = \frac{950}{l} * 0.001$ Amperes. Therefore, for a given data transfer that lasts t seconds, we must have the energy consumption $E(d)$ obtained as the sum of instantaneous power as follows:

$$E(d) = \int_0^t V_c * I \Delta t \text{ Wh} \tag{3}$$

$$= \int_0^t V_c * \left(\frac{950}{l}\right) * 0.001 \Delta t \text{ Wh} \tag{4}$$

$$E(d) = V_c * \left(\frac{950}{l}\right) * 0.001 * t \text{ Wh} \tag{5}$$

3.2 Experiments

In this study we use a Nokia N95 smart phone, which runs Symbian OS version 9.2 with S60 3rd Edition. The experiments for network access were performed using a php/perl server and python scripts as clients on the phone. The device has an idle battery lifetime of 230 hours and a battery capacity of 950 mAh. While we restrict our experiments to a single phone for consistency purposes, we believe the relationships we establish will hold for other devices with similar battery capacities and network configurations.

We conducted data transfer experiments using both the cellular data network GSM and Wi-Fi network interfaces. Each data transfer experiment proceeded thus: First, the battery is charged until it reaches its full capacity. Then a client application is invoked on the phone which issues an http request to a server. The server upon receiving the request transmits a fixed amount of data, some N KB, to the phone. The phone upon completing the download, waits for a fixed interval of time, 10 seconds, and then issues another http request. The whole process is repeated until the battery drains completely. For all the experiments, for each request/response, the data transfer time (including the time to issue a request and receive a response), the amount of data received and the system level power indicator measurement were logged upon completion of each data transfer on the phone.

3.3 Measurements based Model

Figure 2 shows the energy consumption and time taken for data transfers of size 1K to 500K over GSM. The plot shows that the energy consumption has a linear relationship with the data size – for the data sizes we consider, increase in size of data causes a proportional increase in the energy consumption.

From the plots, we infer two linear models to predict the energy consumed and the average time taken for transferring data of a given size over GSM. The energy and time consumed by transferring data of size n kilobytes are given by the following equations:

$$E(n) = 0.0086(n) + 0.31 \text{ Wh} \tag{6}$$

$$t(n) = 0.02(n) + 2.01 \text{ seconds} \tag{7}$$

Plugging the data transfer energy model back into Equation 1 the energy consumed per query is

$$E(q) = \sum_{d \in D(q)} (0.0086(|d|) + 0.31) \text{ Wh} \tag{8}$$

The linear energy model is consistent with previous studies of energy consumption of GSM/EDGE [2]. There are two characteristics of the energy model that are of interest to us: First, both reducing network activity in terms of number of transfers and the data size per transfer result in energy savings. Second, from the two models we see that there is a large overhead associated with each data transfer for both energy and time taken to transfer. This suggests that reducing the data transfers at the cost of increasing the data size per transfer can lead to an overall reduction in energy consumption without incurring an increase in latency.

In the subsequent sections we explore two ideas to exploit these opportunities for saving energy: 1] In the standard IR model we attempt to reduce the overall amount of network activity by improving effectiveness of the IR system. 2] To reduce the number of transfers, we look at modifications to the interaction model that results in bundling more data at each transfer opportunity.

4. EFFECTIVENESS AND ENERGY

PR@10	Total Data (MB)	Energy/Query
0.1	137.96	1.95
0.5	55	0.80
0.9	45.7	0.65

Table 2: Retrieval Effectiveness vs. Energy Consumption: PR@10 is the precision at the rank where the 10 relevant document was retrieved.

Improving retrieval effectiveness reduces the overall network activity in two ways: 1] by reducing the overall amount of data that gets transferred and 2] by reducing the number of transfers. To study the impact of retrieval effectiveness on energy consumption, we simulated three systems with varying levels of retrieval effectiveness measured by the precision at the rank where the 10th relevant document was retrieved, denoted by *PR@10* (not to be confused with P@10). For example, with a PR@10 of 0.5, the 10th relevant document would be retrieved at rank 20. To illustrate the potential savings, we assume that for each query, we know the location of the 10th relevant document in the rank list and that the user will not be interested in perusing the rank list beyond the 100th document. We conducted three retrieval experiments on the phone corresponding to each level of system effectiveness. Each experiment proceeded thus: For each query, we transfer a bundle of documents created by adding relevant documents and non-relevant documents to obtain the desired precision level. Then, we wait for a fixed interval of time and then repeat the process for the next query until the battery drains completely. Then, we can obtain the average energy per query using Equation 5. Table 2 displays the measurements using a set of 50 queries from the TREC Robust 04 collection.

Increasing system effectiveness decreases the overall amount of data transferred and results in reduced energy consumption. Interestingly, the drop in energy consumption is higher for improvements from 0.1 to 0.5 than from 0.5 to 0.9. Figure 3 shows that in the low effectiveness region small improvements lead to larger reductions in data transferred explaining the corresponding observations in the energy measurements.

To observe differences due to smaller improvements in system effectiveness we analyze the energy consumption using the energy

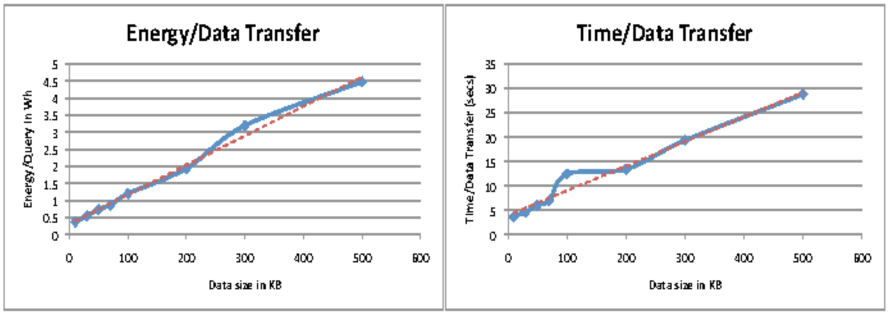

Figure 2: Energy consumption and transfer times for data transfer of different sizes.

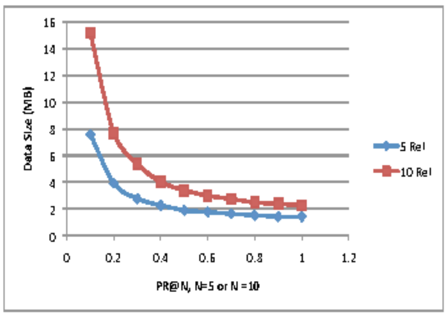

Figure 3: Average response size summed over 50 queries – in MB (y-axis) against the effectiveness of the retrieval system (x-axis) when the user downloads documents until 5 or 10 relevant documents.

model(Equation 8) in the following scenario: For each query the user peruses a ranked list of snippets, 5 at a time, and requests relevant documents until 5 relevant documents are seen. Figure 4 plots energy consumed for queries with different effectiveness. Poorly performing queries cause high energy consumption and the energy benefits of improving precision quickly tapers off confirming the trends we observed on the phone. This indicates that for a reasonably effective retrieval system, we need significant improvements in retrieval effectiveness to obtain substantial energy reductions while for a retrieval system with low effectiveness even small improvements can prove to be useful.

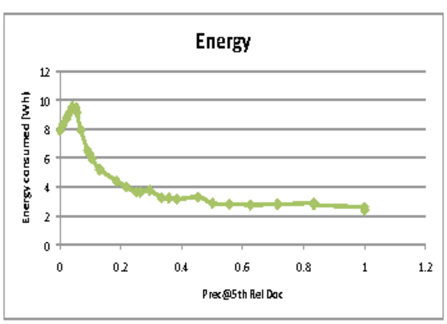

Figure 4: Per Query Energy Consumption for Robust 04: Energy consumption versus effectiveness.

4.1 Relevance Feedback for Reducing Energy

User interaction in the form of relevance feedback has been shown to provide significant gains in retrieval effectiveness. We performed simulations of a relevance feedback interaction on the phone as follows: For each query, the user is presented with a result page of snippets, the user peruses the snippets and downloads relevant documents. When the user requests the subsequent result page we make use of the documents that the user had downloaded to perform relevance feedback. If the user did not download any new documents before requesting the result page then we continue sending the next snippets from the previous retrieval. The process is repeated until the desired number of relevant documents is obtained or up to 100 snippets have been examined. Figure 5 displays energy consumption using the baseline retrieval method and the relevance feedback based method using the in the TREC Robust 04 collection for varying number of required relevant documents. When retrieving 5 relevant documents the feedback method has lower energy consumption compared to the baseline method however as the number of relevant documents pursued by the user increases we see that the feedback method's energy consumption increases. We observed that in this dataset, the size of the relevant documents retrieved by the feedback method increased significantly when compared to the baseline method which explains the reversal of energy gains.

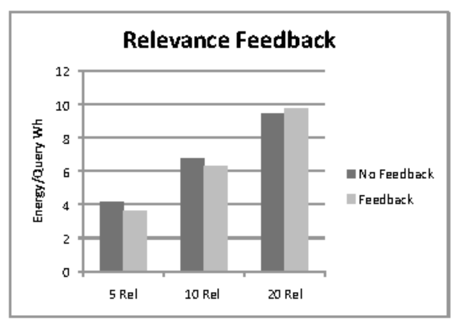

Figure 5: Energy consumption (y-axis) of standard retrieval and relevance feedback retrieval when the number of relevant documents downloaded are varied (x-axis).

Improving system effectiveness reduces the overall amount of data transferred, but the resulting energy savings are limited, particularly when the baseline system accuracy is reasonable. Furthermore, only using positive relevance feedback does not provide much improvement for poorly performing queries, further limiting the opportunity for reducing energy consumption using feedback. Figure 6 shows the average number of result pages for each query under different levels of system effectiveness. We see that the reduction in the number of result pages and the number of data transfers is small even with large improvements in system effectiveness. This result suggests that we should consider ways that directly reduce the number of transfers.

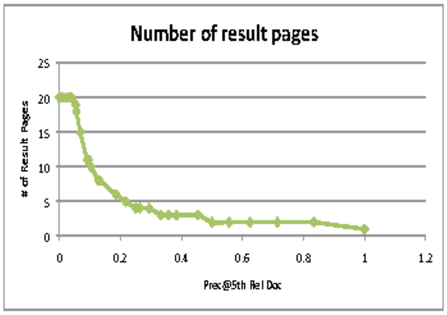

Figure 6: Effectiveness versus Number of Result Pages.

5. INTERACTION MODELS

In this section, we focus on the interaction models with the goal of reducing the number of data transfers per query. Typically, the number of data transfers for a given query corresponds to the sum of the number of result pages that the user looks at and the number of documents that the user downloads. The number of result pages depends on the interaction model used by the IR application. For the rest of this section, we will term the data that gets transferred in each transfer opportunity as a *bundle*.

5.1 Standard

In the standard IR interaction model, for each query, the ranked list is divided into some number of result pages, typically containing about 10 snippets each. Thus, each bundle corresponds to a result page or a relevant document from a result page requested by the user.

5.2 Bundling

There are two ways to reduce the number of data transfers through *bundling* – sending more data at a transfer opportunity: 1] Increase the number of results shown per result page; 2] In addition to sending the snippets, send the documents corresponding to the snippets at each transfer opportunity. The former reduces the number of result pages that get transferred whereas the latter also reduces data transfers needed to transfer documents. Obviously, bundling increases the amount of data that gets transferred in each data transfer and in the second case, bundling also increases the total amount of data that gets transferred. However, based on the energy model we expect that this increase in the data size per bundle can be offset by the resulting reduction in number of transfers.

While the large constant in the energy model allows the increases in data sizes to be offset by reductions in number of data transfers, the actual gains are largely dependent on the actual increase in data sizes and the likelihood of reducing data transfers. Therefore, we use two simple heuristics to avoid increasing the data sizes too much and to avoid sending data that is not likely to be useful. First, we avoid adding documents that can be exceptionally long compared to the average document size. This reduces the risk of increasing the data size too much compared to the potential savings we can expect. Second, the likelihood of reducing data transfers by bundling documents reduces with increase in rank. It is well known and countless empirical experiments have demonstrated that, in almost all cases, the likelihood of finding relevant documents in the result list falls exponentially with the rank. Therefore, we modify the bundling approach to be more selective by only sending documents for the top N ranks. While we tried other approaches based on the estimated precision of the rank list, we found a simple empirically determined rank (15) turned out to be as effective.

5.3 Streaming

We extend the bundling approach further by sending one bundle that contains all the snippets and documents retrieved for the query (the top N). This approach reduces the number of transfers to one. For purposes of comparison with bundling, we stream the data as follows: We stream the first b snippets in the ranked list, followed by the documents corresponding to those snippets, followed by the next b snippets in the ranked list followed by the corresponding documents. If the user requests a particular item from the stream and the item is not already available, the current stream is abandoned and streaming starts from the requested item. Also, to keep the analysis amenable, we assume that the user spends a fixed amount of time on each relevant document and then proceeds to make the next request.

6. MODEL BASED EVALUATION

In this section we present the model based evaluation of the proposed interaction models. Specifically, we show the following: 1] the energy impact of bundling and streaming, 2] the energy impact of feedback on the interaction models, and 3] the effect of bundling on effectiveness due to feedback.

We analyze the energy consumption of the three interaction models based on the energy model (Equation 8) when the user is interested in retrieving 5 relevant documents.

- *Standard* - Each bundle consists of a fixed number of snippets. User requests for documents are served by individual transfers.
- *Bundling* - Each bundle consists of a fixed number of snippets and also the documents that correspond to them.
- *Selective Bundling* - Each bundle consists of a fixed number of snippets. For the bundles that contain the top 15 ranks, the documents are also included.
- *Streaming* - A single bundle consisting of the entire result list is streamed. User request for documents or snippets page results that are not already available results in the stream switching to the requested item.

As in previous analyses we will assume that the user only looks at relevant documents and will terminate perusing the ranked list beyond 100 documents.

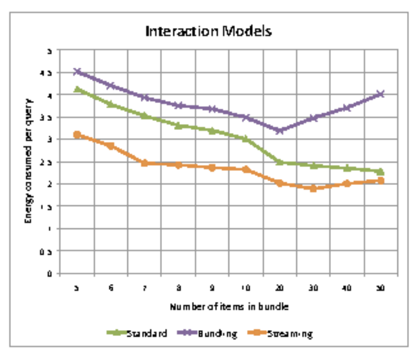

Figure 7: Aggressive Bundling: Energy consumption according to the energy model.

6.1 Bundling and Streaming

Figure 7 shows the energy consumption behavior for the three interaction models according to the energy model.

Standard model shows consistent reduction in energy with increase in the number of snippets shown per page due to fewer result pages that get transferred to obtain the desired number of relevant documents.

Bundling model consumes more energy than the *standard* model. A closer per query analysis in Figure 8 shows the relationship between retrieval effectiveness versus energy consumption under the different interaction models. As we had seen earlier, for poorly performing queries more data is retrieved using more data transfers thereby causing increased energy consumption for all the interaction models. However, the energy consumption for the bundling and streaming models are much higher than the standard model for poorly performing queries. The energy saved by reducing the number of transfers is much less compared to the additional energy spent in transferring the documents, most of which are not relevant. However, the bundling and streaming models provide substantial savings in energy for queries with high precision due to the reduction in number of transfers and the relatively low energy cost incurred in transferring documents. But mostly, the energy reductions obtained for queries with higher precision are lost on poor queries.

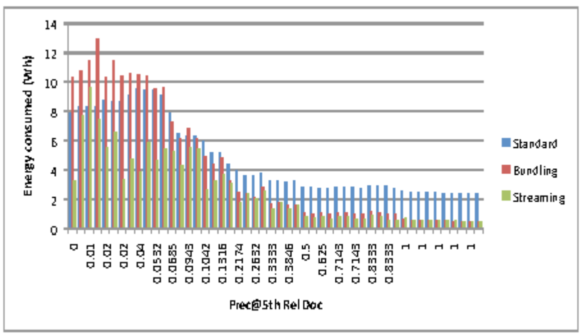

Figure 8: Effectiveness versus Energy: Plot of energy consumption of queries against their precision at 5^{th} relevant document.

Selective bundling model consistently outperforms the *Standard* model for all bundle sizes. By taking a selective approach to bundling, this model benefits from reducing the number of transfers that correspond to the document requests while reducing the number of unnecessary data transfers. The results we report are for a simple rank based threshold (15) approach which we were able to easily implement on the phone. We tried other adjustments such as avoiding bundling large documents and dynamically selecting the rank threshold using an estimate of the precision of the ranked list. We were able to observe an additional 5% improvement over the results that we report here.

Streaming model provides the best savings in energy compared to all the other interaction models, but the benefits of the model are largely dependent on the user activity. We assume that after a fixed perusal time the user will switch to the next data request. However, the time a user spend on a document is hard to model. Nevertheless, the results indicate that, when possible, a data stream that is dynamically adjusted based on user interactions can reduce energy consumption by reducing the number of transfers.

6.2 Bundling with Feedback

Figure 11 shows energy consumption for the interaction models with user feedback. In the aggressive *bundling* model, relevance feedback reduces the number of bundles that get transferred and even saving a single bundle can result in significant savings in energy. However, overall the benefits of feedback do not outweigh the costs incurred by bundling documents in all the result pages. On the other hand, the benefits are more pronounced for the *selective bundling* model. In addition to reducing the number of bundles transferred due to increase in overall precision (PR@5 rose from 0.44 to 0.51), relevance feedback also improves the likelihood of relevant documents in the first few bundles (the precision @ rank 10 rose from 0.41 to 0.47) thus increasing the number of relevant documents in the initial bundles. Finally, we also observe that the *Streaming* model benefits substantially from the reduction in the size of the stream due to feedback. Thus, while relevance feedback provided limited gains for standard IR, the *bundling*, *selective bundling*, and *streaming* models gain substantial benefits from relevance feedback.

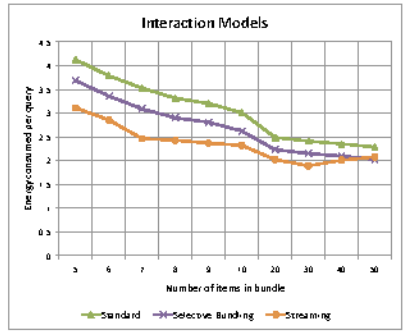

Figure 9: Selective Bundling: Energy consumption according to the energy model.

6.3 Bundling versus Effectiveness

While bundling reduces energy consumption it introduces an interesting trade-off in effectiveness when performing incremental feedback. For instance, consider the scenario where each result page (bundle) consists of 5 documents – after perusing the snippets and relevant documents (if any) in the result page, the user selects the documents that are relevant before requesting the next page. This feedback is used to return a new result page. When we plotted the effectiveness in terms of the precision at the 5th relevant document, we find that (see Figure 10) the effectiveness drops with the increase in the bundle size. This is because with increase in the number of results in a page we obtain feedback less often and display more results to the user.

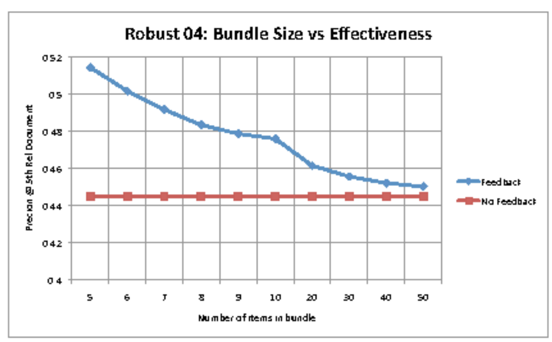

Figure 10: Bundle Size versus Effectiveness.

7. IMPLEMENTATION RESULTS

To validate the energy model based findings, we performed actual retrieval experiments from a mobile phone. We used a mock-user agent that simulates the actions such as issuing queries, perusing snippets and downloading relevant documents. Figure 12 shows the results of actual runs on a mobile phone under the dif-

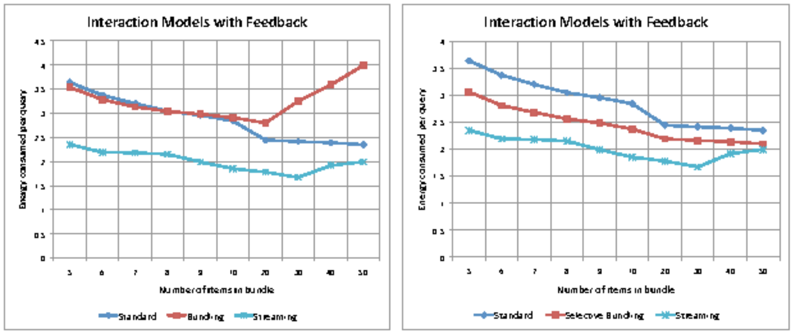

Figure 11: Relevance Feedback for Bundling: Energy consumption according to the energy model. Figure on the left is for retrieval with feedback using selective bundling and the one on the right is for retrieval with feedback using selective bundling.

ferent interaction models. We note that the retrieval experiments have a higher frequency of data transfers compared to the experiments used to build the energy model. This explains the slightly higher rate of energy conumption in the implementation experiments compared to the model-based experiments. However, the relative difference in performance observed in the implementation experiments are consistent with our model.

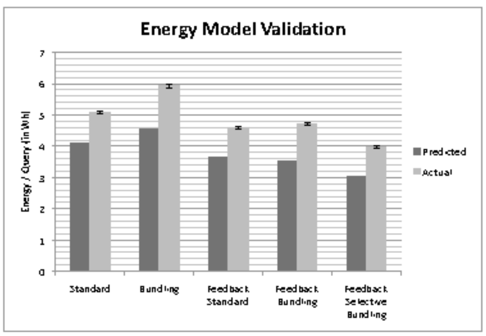

Figure 12: Predicted and actual energy consumption rates observed on the phone.

The results confirm our model based findings and also demonstrate the predictive accuracy of the energy model for different retrieval experiments. Energy measurements on the phone are affected by various factors including the time at which the experiments were conducted, the location, signal strength and others. For instance, we observed consistently higher energy consumption and higher energy overhead for measurements made during the day when the data network channel is likely to be more congested. The measured improvements of the selective bundling with feedback over the standard IR model during the day rose to 22% from 18% during the night. Similarly, energy measurements of data transfers are higher when conducted from low signal strength locations. The measurements used in this study were controlled for consistency in terms of the location, the time of the day, the device used and the battery used. We conducted two experiments on another Nokia N95 device with a different battery but with the same specifications and found the results to be consistent with the one used in this paper.

Dead time – time spent in waiting for data to download – and the overall *latency* to access a fixed number of relevant documents are important considerations for the interaction models. As with energy, we expect an increase in the amount of data transferred to result in increases in the transfer times. However, we find that as was the case with energy, the *bundling* model, results in improvements in both metrics compared to the *Standard* model. Figure 13 shows the measured average dead time and latency for the different methods based on actual runs on the mobile phone. The selective feedback approach saves around 10 seconds in dead time per query on average and around 13 seconds in overall latency considering consistent user behavior across the different runs.

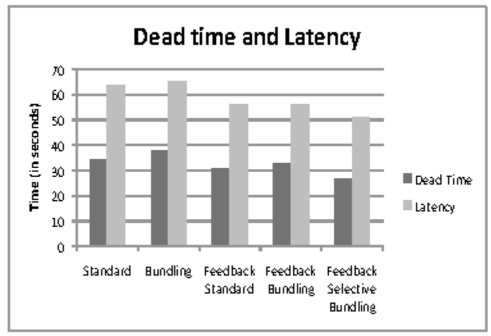

Figure 13: Deadtime and latency measurements on actual runs of the implementation of different methods on the mobile phone.

To summarize, we observe that for the data sizes of the Robust 04 collection, altering the interaction models leads to reduced energy consumption by reducing the number of data transfers at the cost of increasing the size of each individual data transfer. The predicted energy savings of the selective bundling model with feedback ranges up to 25% and the observed savings around 18% compared to the standard IR model with no feedback and the results translated into corresponding increases in number of queries handled. While relevance feedback did not reduce energy, we see that it helps improve the impact of the bundling approaches. Furthermore, when the number of required relevant documents increases we observe that energy reduction due to the bundling model increases consistently as shown in Figure 14. Finally, our simple approach to building a reliable energy model is easily repeatable on devices whose remaining battery life accuracy and granularity are not adequate.

8. COLLECTION CHARACTERISTICS

There are two main characteristics of the Robust 04 collection that benefit the bundling models. First, the document sizes are relatively small compared to the size of the snippets. Therefore, sending documents instead of snippets for a few documents does not dramatically increase the energy consumption compared to energy savings obtained by reducing few data transfers. Second, the queries for Robust 04 collection contain several relevant documents and it is fair to assume that users interested in searching this collection would look for many relevant documents for each query.

However, these assumptions do not hold on larger web type document collections where the average document size (around 25K)

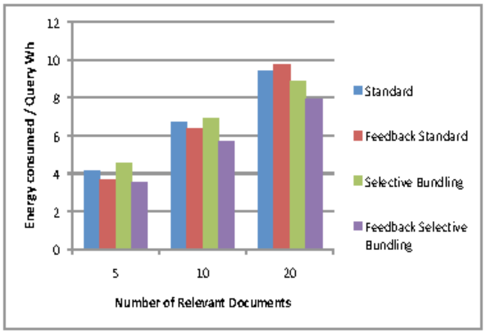

Figure 14: Energy consumption of the different interaction models for increasing number of relevant documents downloaded by the user.

is much higher and a large fraction of the queries have few relevant documents (e.g., home page and navigational queries). To evaluate the benefits of the models on larger web type document collections we ran the interaction models on the TREC Gov2 collection consisting of 25 million web pages from the .gov domain.. The results for the interaction models are shown in Figure 15. The energy reductions we observe for this collection are modest compared to the Robust 04 results. The increase in data sizes far outweigh the reductions obtained through reducing the number of transfers in the *bundling* models. Furthermore, the streaming model consumed more energy compared to the *Standard* model, as most of the data the streaming model takes longer to deliver documents and is never able to stream enough data ahead of time.

Certain extensions can be implemented to further reduce the energy consumption for web documents. For example, the impact of bundling can be improved further with prior information on the effectiveness of queries and the prior probability of clicks on documents in the result set. These extensions are part of future work.

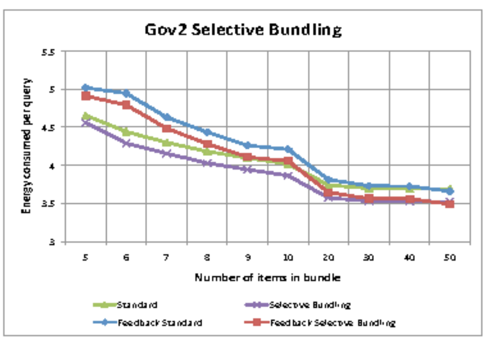

Figure 15: Gov2 Collection: Energy consumption of the different interaction models.

9. CONCLUSIONS

In this work, we studied energy consumption of network based information retrieval resulting in the following contributions:

- We characterized the energy consumption of an IR application and described an energy model that can be used to analyze energy consumption of different retrieval and interaction models.
- Based on the energy model, we found that systems with low effectiveness have high energy consumption. Across all effectiveness levels, we need large improvements in effectiveness in order to observe substantial energy savings, but that is particularly true when a system's baseline performance is good.
- Our results show that for the energy model we observe, adjusting the interaction models to reduce the number of transfers can result in substantial energy savings with improvements in *dead time* and *latency*.

While our work was limited to the GSM/EDGE network, we can extend this analysis to the more advanced 3G cellular network. 3G networks have higher bandwidth but also higher energy consumption rates. In our analyses, we have assumed that the user only clicks on relevant documents. However, we have not considered the effect of display constraints that limit the space available for snippets, in turn lowering the accuracy of perceived relevance. In such scenarios bundling based models can be useful. Search is often an entry point for browsing information on the web. A natural extension of our work is to analyze the energy consumption of browsing behavior originating from search. In conclusion, our work provides a starting point for the analysis of energy costs of network based retrieval applications.

10. ACKNOWLEDGMENTS

This work was supported in part by the Center for Intelligent Information Retrieval. Any opinions, findings and conclusions or recommendations expressed in this material are the authors' and do not necessarily reflect those of the sponsor.

11. REFERENCES

[1] D. Arter, G. Buchanan, M. Jones, and R. Harper. Incidental information and mobile search. In *MobileHCI '07: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, pages 413–420, New York, NY, USA, 2007. ACM.

[2] D. Charlet, V. Issarny, and R. Chibout. Energy-efficient middleware-layer multi-radio networking: An assessment in the area of service discovery. *Comput. Netw.*, 52(1):4–24, 2008.

[3] K. Church and B. Smyth. Who, what, where & when: a new approach to mobile search. In *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*, pages 309–312, New York, NY, USA, 2008. ACM.

[4] K. Church, B. Smyth, P. Cotter, and K. Bradley. Mobile information access: A study of emerging search behavior on the mobile internet. *ACM Trans. Web*, 1(1):4, 2007.

[5] S. Gaonkar, J. Li, R. R. Choudhury, L. Cox, and A. Schmidt. Micro-blog: sharing and querying content through mobile phones and social participation. In *MobiSys '08: Proceeding of the 6th international conference on Mobile systems, applications, and services*, pages 174–186, New York, NY, USA, 2008. ACM.

[6] D. Huang, A. Trotman, and S. Geva. Efficient Information Seeking on a Mobile Device. In *SIGIR' 08: Proceedings of the Workshop on Mobile Information Retrieval*, 2008.

[7] M. Jones, G. Buchanan, R. Harper, and P.-L. Xech. Questions not answers: a novel mobile search technique. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 155–158, New York, NY, USA, 2007. ACM.

[8] M. Kamvar and S. Baluja. A large scale study of wireless search behavior: Google mobile search. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 701–709, New York, NY, USA, 2006. ACM.

[9] E. W. D. Luca and A. Nürnberger. Supporting information retrieval on mobile devices. In *MobileHCI '05: Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, pages 347–348, New York, NY, USA, 2005. ACM.

[10] J. Otterbacher, D. Radev, and O. Kareem. Hierarchical summarization for delivering information to mobile devices. *Inf. Process. Manage.*, 44(2):931–947, 2008.

[11] A. Rahmati and L. Zhong. Context-for-wireless: context-sensitive energy-efficient wireless data transfer. In *MobiSys '07: Proceedings*

of the 5th international conference on Mobile systems, applications and services, pages 165–178, New York, NY, USA, 2007. ACM.

[12] K. Rodden, N. Milic-frayling, R. Sommerer, and A. Blackwell. Effective web searching on mobile devices. In *In Proc. HCI*, pages 281–296. Springer Verlag, 2003.