Puppeteer: Component-based Adaptation for Mobile Computing

Eyal de Lara[†], Dan S. Wallach[‡], and Willy Zwaenepoel[‡]

[†] Department of Electrical and Computer Engineering
[‡] Department of Computer Science
Rice University

Abstract

Puppeteer is a system for adapting component-based applications in mobile environments. Puppeteer takes advantage of the exported interfaces of these applications to perform adaptation *without* modifying the applications. The system is structured in a modular fashion, allowing easy addition of new applications and adaptation policies.

Our initial prototype focuses on adaptation to limited bandwidth. It runs on Windows NT, and includes support for a variety of adaptation policies for Microsoft PowerPoint and Internet Explorer 5. We demonstrate that Puppeteer can support complex policies without any modification to the application and with little overhead. To the best of our knowledge, previous implementations of adaptations of this nature have relied on modifying the application.

1 Introduction

The need for application adaptation in mobile and wireless environments is well established [6, 11, 12, 19, 30, 31]. Many approaches to adaptation have been proposed before, and many taxonomies of adaptation are possible. We focus here on where the adaptation is implemented, and recognize system-based [20, 25] and application-based adaptation [13, 14, 17, 32] as the opposite ends of the spectrum. As the names imply, with system-based adaptation, all adaptation is done once and for all in the system, while with application-based adaptation, each application implements its own adaptations. Application-based adaptation enables a much larger class of adaptations, but requires modification of the applications.

In this paper, we present a novel approach to adaptation we call *component-based adaptation*, which enables application-specific adaptation policies *without* requiring modifications to the application. The novel feature of component-based adaptation is that it uses the exposed APIs of component-based applications to implement application-specific policies. Component-based adaptation attempts to bring together the benefits of system-based and application-based adaptation, namely to implement application-specific policies without modifying the applications.

Component-based adaptation is by nature restricted to component-based applications with exported APIs. While certainly a limitation, we observe that many desirable candidate applications for adaptation are already component-based, including the Microsoft Office Suite, Internet Explorer, Netscape Navigator, the KDE Office Suite, and Star Office. Recognizing the advantages of component-oriented software construction – independent of adaptation – we foresee an increasing number of such applications being developed as components with exported APIs. Although traditionally associated with the Windows platform and with COM/DCOM technology, component-based technologies are becoming more common in the UNIX world as well, where the push for component-based technologies is led by the GNOME [1] and KDE [3] projects. These open source initiatives strive to enable the development of applications that allow seamless integration of components and have a consistent look-and-feel. A good example is KOffice [4], an open source productivity suite with powerful scripting capabilities. More recently, StarOffice [5] released version 5.2 of its popular cross-platform productivity suite, which implements a sophisticated object model that allows scripting by third party applications through a CORBA-based interface.

The more fundamental question about component-based adaptation is to what extent it can support the adaptation mechanisms that a customized application-based approach can achieve and with what performance. Furthermore, we wish to understand the scalability of component-based adaptation. Clearly, the system needs "drivers" for each application it wishes to support. For the concept to be scalable in terms of the number of applications it supports, the effort involved in writing an additional driver must be made small.

To address these questions, we have built a system we call Puppeteer. This paper describes the design of the Puppeteer system, its implementation on Windows NT, and our experience using this implementation to adapt two applications – Microsoft PowerPoint (a presentation graphics system, hereafter "PowerPoint") and Internet Explorer

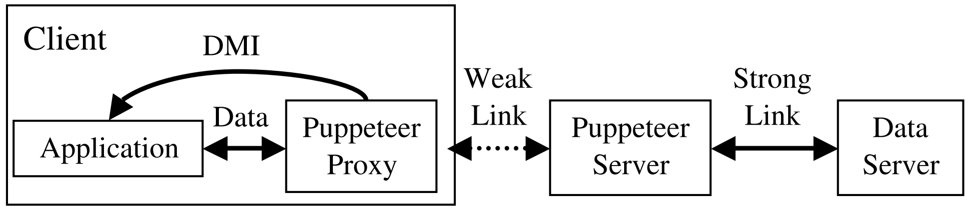

Figure 1: System architecture.

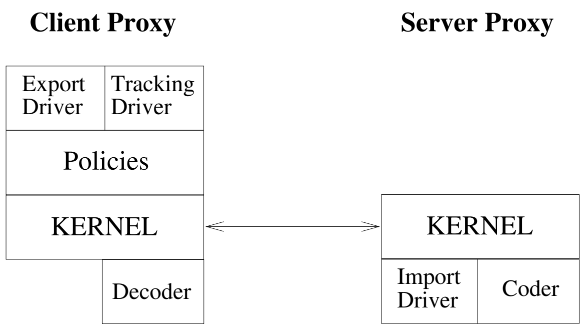

Figure 2: Puppeteer architecture.

5 (a Web browser, hereafter "IE") – for low bandwidths. We demonstrate that Puppeteer can easily and efficiently support a number of desirable policies. We are currently also in the process of porting Puppeteer to Linux and expect to support adaptation of StarOffice applications in the near feature.

The rest of this paper is organized as follows. Section 2 presents the architecture of the Puppeteer system. Section 3 introduces the applications we use to evaluate the prototype. Section 4 describes the experimental platform. Section 5 describes the documents we use in our experiments. Section 6 presents our experimental results. Section 7 discusses related work. Finally, Section 8 discusses our conclusions.

2 Puppeteer

Figure 1 shows the four-tier Puppeteer system architecture. It consists of the application(s) to be adapted, the Puppeteer client proxy, the Puppeteer server proxy, and the data server. The application and data server are completely unmodified. The Puppeteer client proxy and server proxy work together to perform the adaptation.

The Puppeteer client proxy is in charge of executing the policies that adapt the applications. The Puppeteer server proxy is assumed to have strong connectivity to the data server. In the most common scenario, it executes on the same machine as the data server. The Puppeteer server proxy is responsible for parsing documents, exposing their structure, and transcoding components as requested by the client proxy. Data servers can be arbitrary repositories of data such as Web servers, file servers or data bases.

2.1 Application Requirements

Puppeteer can adapt an application if it can uncover the component structure of its documents and if the application provides a run-time interface that enables Puppeteer to view and modify the data the application operates on. We refer to the latter feature as Data Manipulation Interface (*DMI*). Additionally, Puppeteer can benefit greatly from being able to track the user as she uses the application. We demonstrate next how the Puppeteer implements adaptation once these requirements are met.

2.2 Puppeteer Architecture

The Puppeteer architecture consists of four types of modules: Kernel, Driver, Transcoder, and Policy (see Figure 2). The Kernel appears once in both the client and server Puppeteer proxy. A driver supports adaptation for a particular component type. A driver for a particular component type may call on a driver for another component type, if a component of the latter type is included in a component of the former type. At the top of this driver hierarchy sits the driver for a particular application (which itself is a component type). Drivers may execute both in the client and the server Puppeteer, as may Transcoders which implement specific transformations on component types. Policies specify particular adaptation strategies and execute in the client Puppeteer proxy.

2.2.1 Kernel

The Kernel is a component-independent module that implements the Puppeteer protocol. The Kernel runs in both the client and server proxies and enables the transfer of document components. The Kernel does not have knowledge about the specifics of the documents being adapted. It operates on a format-neutral description of the documents, which we refer to as the Puppeteer Intermediate Format (PIF). A PIF consists of a *skeleton* and a set of *data items*. The skeleton captures the structure of the data used by the application. The skeleton has the form of a tree, with the root being the document, and the children being pages, slides or any other elements in the document. The skeleton is a multi-level data structure as components in any level can contain sub-components. The data items contain the native data of individual components. A component can have one or more related data items.

When adapting a document, the Kernel first communicates the skeleton between the server and the client proxy. It then enables application policies to request a subset of the components and to specify transcoding filters to apply to the component's data.

2.2.2 Drivers

For every component type it adapts, Puppeteer requires an import and an export driver. To implement complex policies, a tracking driver is also necessary. The import drivers parse the documents, extracting their component structure and converting them from their application specific file formats to PIF.

Most import drivers construct the skeleton by statically parsing the document in the Puppeteer server proxy. When the application only exposes a DMI, but has an opaque file format, Puppeteer runs an instance of the application on the server, and uses the DMI to uncover the structure of the data, in some sense using the application as a parser. This configuration allows for a high degree of flexibility and makes porting applications to Puppeteer more straightforward as Puppeteer need not understand the application's file format. It creates, however, a large overhead on the server proxy, resulting in lower performance. Moreover, it requires both the client and server to run the environment of the application, which in most cases amounts to having to run the same operating system in both servers and clients. In contrast, when the file format is parsable, either because it is human readable (*e.g.*, XML) or there is sufficient documentation to write a parser, Puppeteer can parse the file(s) directly to uncover the structure of the data. This results in better throughput and enables clients and server to run on different platforms (*e.g.*, running the Puppeteer client proxy on Windows NT while running the Puppeteer server proxy on Linux).

However, static parsing at the server does not work well for documents that choose what data to fetch and display by executing a script, or by other dynamic mechanisms. Instead, import drivers for dynamic content run on the Puppeteer client proxy and rely on an intercept mechanism that traces requests.

Regardless of whether the skeleton is built statically in the server or dynamically in the client proxy, any changes to the skeleton are reflected by the Kernel at both ends to maintain a consistent view of the skeleton.

Export drivers un-parse the PIF and update the application using the DMI interfaces exposed by the application. A minimal export driver, has to support inserting new components into a running application.

Tracking drivers are necessary for many complex policies. A tracking driver tracks which components are being viewed by the user and intercepts load and save requests. Tracking drivers can be implemented using polling or event registration mechanisms.

2.2.3 Transcoders

Puppeteer makes extensive use of transcoding to perform transformations on component data. Transcoders include the conventional ones, such as compression and reducing image resolution. A novel transcoding mechanism is used to enable loading subsets of components. Each element of the PIF skeleton has a number of associated data items that, among other things, encode in a component-specific format the relationship between the component and its children. To load a subset of the children of a given node, it is sometimes necessary to modify the data items associated with the parent node to reflect the fact that we are only loading some of its children. In effect, by transcoding the parent node's data items, we create a new temporary component that consists only of a subset of the children of the original component.

2.2.4 Policies

Policies are modules that run on the client proxy and control the fetching of components. These policies traverse the skeleton, choosing what components to fetch and with what fidelity.

Puppeteer provides support for two types of policies: general purpose policies that are independent of the component type being adapted (*e.g*, prefetching); and component specific policies that use their knowledge about the component to drive the adaptation (*e.g.*, fetch the first page only) .

Typical policies choose components and fidelities based on available bandwidth and user-specified preferences (*e.g.*, pre-fetch all text first). Other policies track the user as she runs the application and try to anticipate her needs (*e.g.*, fetch the PowerPoint slide that currently has the user's focus and pre-fetch subsequent slides in the presentation), or react to the way the user moves through the document (*e.g.*, if the user skips pages, the the policy can drop components it was fetching and focus the available bandwidth on fetching components that will be visible to the user).

Regardless of whether the decision to fetch a component is made by a general purpose policy or by a component specific one, the actual transcoding and transfer of component's data is performed by the Kernel, relieving the policy from the intricacies of communication.

2.3 The Adaptation Process

The adaptation process in Puppeteer is divided roughly into three stages: parsing the document to uncover the structure of the data, fetching selected components at specific fidelity levels, and updating the application with the newly fetched data.

When the user opens a (static) document, the Kernel on the Puppeteer server proxy instantiates an import driver for the appropriate document type. The import driver parses the document, extracts its skeleton and data, and generates a PIF. The Kernel then transfers the document's skeleton to the Puppeteer client proxy. The policies running on the client proxy policy ask the Kernel to fetch an initial set of components from within the skeleton at a specified fidelity.

The policies then use the export driver to supply this set of components to the application as though it had the full document at its highest level of fidelity. The application, believing that it has finished loading the document, returns control to the user. Meanwhile, Puppeteer knows that only a fraction of the document has been loaded and will use the techniques described above to fetch or upgrade the fidelity of the remaining components.

3 Prototype

We chose to support PowerPoint and IE as the first two applications for our initial prototype. Besides being widely popular, these two applications comply with all or most of the requirements for DMI, parsable file formats, and tracking mechanism from Section 2.1. Furthermore, PowerPoint and IE have radically different DMIs. By supporting both, we are more likely to accurately design the interfaces between the Puppeteer Kernel and the component-specific aspects of the system. Next, we discuss the design of the drivers, transcoders, and policies that we implemented to adapt these two applications. Table 1 shows the code line counts for the various modules.

Module		Code Lines
Kernel		7996
PPT	Import Driver	1114
	Export Driver	807
	Track Driver	112
	Transcoders	392
	Policies	287
	Total	2712
IE	Import Driver	314
	Export Driver	347
	Track Driver	65
	Transcoders	749
	Policies	334
	Total	1809

Table 1: Codel line counts for Kernel, PowerPoint (*PPT*) and IE modules .

3.1 Drivers

3.1.1 Import drivers

PowerPoint 2000 supports two native file formats: the traditional binary format based on OLE archives [21, 22], and a new XML-based format [23]. We choose to base the PowerPoint import drivers on the XML representation because it is semantically comparable to the binary format and the human readable nature of XML makes it easier to parse and manipulate the document. We implemented import drivers for the following component types: PowerPoint, Slide, Images, Sound, Embedded Objects.

While HTML is straightforward to parse, the introduction of JavaScript in DHTML [15] has allowed for documents whose structure can change dynamically. For DHTML, the import driver intercepts URL requests from within a page, allowing it to dynamically add new images and components to a Web page's skeleton (see Section 2.2.2). We implemented import drivers for the following component types: IE, Images.

3.1.2 Export drivers

PowerPoint and IE DMIs are based on the Component Object Model (COM) [7] and the Object Linking and Embedding (OLE) [8] standards. The interfaces they provide are reasonably well documented [24, 29] and have traditionally been used to extend the functionality of third party applications.

The PowerPoint and IE DMIs provide excellent access to compose and modify internal data structures. To support the policies we implemented for this paper the PowerPoint export drivers includes support for opening and closing presentation, and inserting slides, images and embedded objects. The IE export driver includes support for navigating to a URL and reloading individual items of a page.

Powerpoint supports a cut-and-paste interface to update a presentation. To paste slides into an active PowerPoint presentation, *active*, the PowerPoint export driver creates a new PowerPoint presentation, *helper*, that consists only of the slides we want to paste in. The update process has two stages. In *Stage 1*, the driver instructs PowerPoint to load *helper*. In *Stage 2*, for every slide in *helper*, the driver copies the slide to the clipboard, pastes it into *active*, and deletes any earlier version of the same slide from *active*.

To update an object in IE, the IE export driver instructs IE to reload only the URL associated with the object.

3.1.3 Tracking drivers

PowerPoint's event notification mechanism is very primitive and encompasses just a handful of large-granularity events like opening or closing of document, making it inadequate for tracking the behavior of the user. The PowerPoint tracking driver relies, instead, on polling the DMI to determine the slide currently being displayed.

The IE tracking driver uses IE's rich event mechanism that allows third-party applications to register call-back functions for a wide range of events. The driver uses this interface to detect when the user types a URL, presses the back or forward buttons, clicks on a link, or moves the mouse over an image. The former events are used to instruct the Kernel to open a new HTML document, while the latter is used by policies to drive image fetching and fidelity refinement (*e.g.*, refine the image currently pointed by the mouse).

3.2 Transcoders

The above policies use the following transcoders:

1 **Slide selector**. Creates a virtual presentation consisting of specific slides.
2 **OLE selector**. PowerPoint stores OLE-based embedded objects in single file. This transcoders creates a new file that contains only a subset of selected OLE-based embedded objects.
3 **Progressive JPEG**. Converts GIF and JPEG images into Progressive JPEG and back to JPEG.
4 **GZIP compressor**. Compresses an uncompresses text and binary data using gzip.

3.3 Policies

This section presents some sample adaptation policies that illustrate the power of component-based adaptation. These policies would be difficult to implement in system-based adaptation, because they affect not only the data used by the application, but also its control flow. Such adaptation policies have, to the best of our knowledge, only been implemented by modifying the application. In Puppeteer, however, they are implemented by using the external APIs. As will be demonstrated in Section 6 these policies also result in significant benefit under limited bandwidth conditions.

1 **PowerPoint: First slide**. Fetch only the components of the first slide at their highest fidelity, and return control to the user. Fetch the rest of the presentation in the background.
2 **PowerPoint: Prefetch text**. Fetch all slides, but leave out any images and embedded objects. Monitor the user and fetch images and embedded objects of the slide that has the focus.
3 **IE: Incremental rendering**. Convert all GIF and JPEG images in a HTML page into Progressive JPEG. Load only the first 1/7 of the image, before returning control to user. Refetch with progressively higher fidelity the image pointed by the mouse.

3.4 Adding New Functionality

To adapt a new application with Puppeteer we need to implement drivers, policies, and transcoders for each new component type that is not currently supported by Puppeteer. For example, to enable MS Word we need to add drivers for the Word component type, but we can reuse the drivers and transcoders for the image and embedded object component types that we implemented for PowerPoint (Table 1).

While the effort in adding new applications and new policies is limited by the modular design of Puppeteer, the lack of standard DMIs, event models, and file formats requires new drivers to be written. Designing such standard interfaces is part of our ongoing research.

4 Experimental environment

Our experimental platform consists of two Pentium III 500 MHz machines running Windows NT 4.0 that communicate via a third PC running the DummyNet network simulator [28]. This setup allows us to control the bandwidth between client and server to emulate various network technologies.

All our experiments access data stored by an Apache 1.3 Web server. For the experiments where we measure the latency of loading the documents using the native application, Apache is the only process running on the server. For the Puppeteer experiments, the Apache server and Puppeteer server proxy run on the same machine.

5 Data sets

We selected the set of PowerPoint documents used in our experiments from a collection of Microsoft Office documents that we characterized earlier [10]. The full collection includes 2167 documents downloaded from 334 Web sites with sizes ranging from 20 KB to 21 MB.

We obtained our HTML documents by re-executing the traces of Web client accesses collected and characterized by Cunha *et. al.* [9]. These traces include access from 2 user groups made during a period of 7 months from

November 1994 through May 1995. These traces have 46,830 unique URLs corresponding to 3026 Web sites. For every URL that we were able to access (many pages had either disappeared or were corrupted), we downloaded the HTML file and any images referenced by them. We did not download any documents linked from these pages. In this manner we acquired 3796 HTML files and 15,329 images, comprising 89 MB of data downloaded from 1009 sites. Documents ranged in size from a few bytes to 773 KB, including images.

Because these data sets are so large, transmitting them at low bandwidth without transcoding would take a prohibitive amount of experimental time. We chose to run our experiments on just 92 PowerPoint documents and 182 HTML documents to limit the running time of tests. In the slowest network configuration the selected sets requires 138 minutes for PowerPoint and 55 minutes for HTML to complete the longest test. For completeness, however, we ran one test over the full sets of both document types over a high bandwidth network, verifying that our selected documents and the full document sets produce similar results.

For our PowerPoint experiments, we selected 92 documents by sorting all documents larger that 32 KB into buckets with sizes increasing by powers of 2. We then randomly selected 10 documents from each bucket. The largest bucket, consisting of documents with sizes greater than 16 MB had only 2 documents. Thus, our experimental set has $9 \times 10 + 2 = 92$ members.

For our Web experiments, we selected 182 HTML documents from the downloaded set by sorting all documents larger that 4 KB into buckets with sizes increasing by powers of 2. We then randomly selected 25 documents from each bucket. The largest bucket, consisting of documents with sizes greater than 512 KB had only 7 documents. Thus, our experimental set has $7 \times 25 + 7 = 182$ members.

6 Experimental results

The fundamental question we want to answer in this section is how much overhead we pay for doing the adaptation outside of the application as opposed to by modifying the application. To answer this question in a definitive way we would need to modify the original applications to add the adaption behavior that we achieve with Puppeteer, and compare its performance to the application running with Puppeteer. This is not possible, since we do not have access to the source code of the applications. Instead, we have designed some experiments to measure the various contributing factors to the Puppeteer overhead.

This overhead consists of two elements: a one-time initial cost, and a continuing cost. The one-time initial cost consists of the CPU time it takes to parse the document to extract its skeleton and the network time to transmit the skeleton and some additional control information. Continuing costs come from the overhead of the various DMI commands used to control the application. We assume that other costs, such as network transmission of application data, transcoding, and rendering times are similar for both implementations.

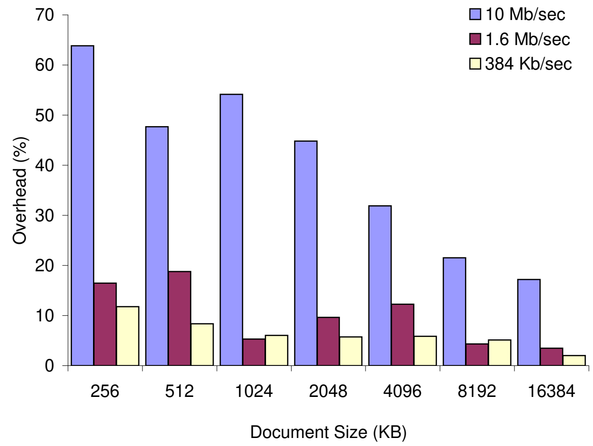

Figure 3: Percentage overhead of PPT.full over PPT.native for various document sizes and bandwidths.

The remainder of this section is organized as follows. First, we measure the one-time initial adaptation costs of Puppeteer. Second, we measure the continuing adaptation costs. Finally, we present several examples of policies that significantly reduce user perceived latency.

6.1 Initial Adaptation Costs

6.1.1 Latency breakdown

To determine the one-time initial costs, we compared the latency of loading PowerPoint and HTML documents in their entirety using the native application (*PPT.native, IE.native*) and the application with Puppeteer support (*PPT.full, IE.original*). In the later configuration, Puppeteer loads the document's skeleton and all its components at their highest fidelity. This policy represents the worst possible case as it incurs the overhead of parsing the document's skeleton but does not benefit from any adaptation.

Figures 3 and 4 show the percentage overhead of PPT.full and IE.original over PPT.native and IE.native for a variety of document sizes and bandwidths. Overall, the Puppeteer overhead varied for PowerPoint documents from 2% for large documents over 384 Kb/sec to 64% for small documents over 10 Mb/sec, and for HTML documents from 4.7% for large documents over 56 Kb/sec. to 305% for small document over 10 Mb/sec.

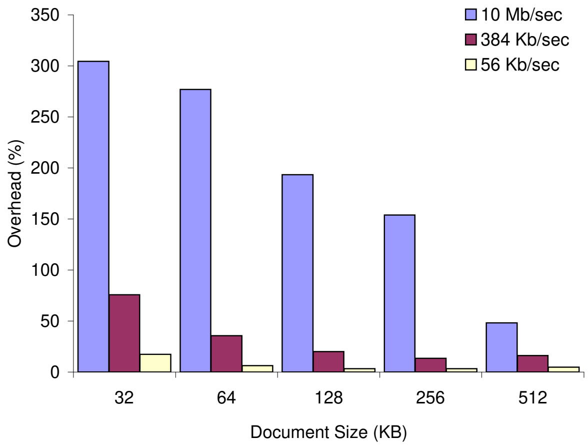

Figure 4: Percentage overhead of IE.original over IE.native for various document sizes and bandwidths.

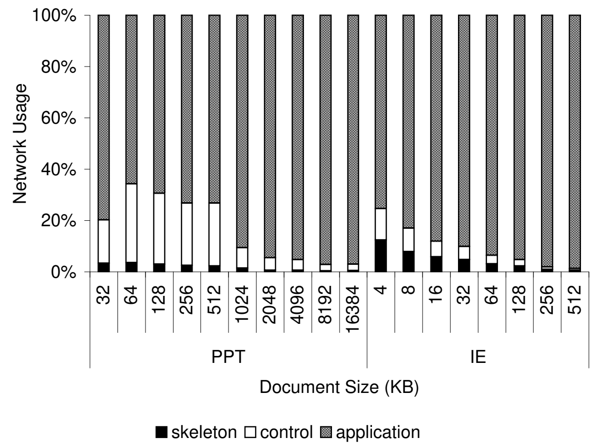

Figure 5: Data breakdowns for loading PowerPoint and HTML documents.

Operation		Cost (ms / component) Single Avg	Single Stdev	Additional Avg	Additional Stdev
Slide (PPT)	Stage 1	746	723	417	492
	Stage 2	148	96	113	99
Image (IE)	Synthetic	N/A	N/A	29	9
	DMI	33	19	32	12

Table 2: Continuing adaptation costs for PowerPoint (*PPT*) slides and IE images. The table shows the cost of executing OLE calls that append PowerPoint slides or upgrade the fidelity of IE images

These results show that the initial adaptation costs of Puppeteer are small compared to the total document loading time, for large documents transmitted over medium and slow network speeds, where adaptation would normally be used.

Figure 5 plots the data breakdown for PowerPoint and HTML documents. We divide the data into application data and Puppeteer overhead, which we further decomposed into data transmitted to fetch the skeleton (*skeleton*) and data transmitted to request components (*control*). This data confirms the results of Figures 3 and 4. The Puppeteer data overhead becomes less significant as document size increases. The data overhead varied for PowerPoint documents from 2.9% on large documents to 34% on small documents, and for HTML documents from 1.3% on large documents to 25% on small documents. This results, however, represent the worst possible case in which the policy fetches components individually. For policies that fetch several components simultaneously, Puppeteer batches control messages, incurring in considerable less overhead.

6.2 Continuing Adaptation Costs

In this section we explore the continuing costs of adapting using the DMI. It is clear that these costs are policy-dependent. Our purpose is not to give a comprehensive analysis of DMI-related adaptation costs, but to show that these costs are small compared to the network and rendering costs, inherent in the application. We perform two sample experiments: loading and pasting newly fetched slides into a PowerPoint presentation, and replacing all the images of an HTML page with higher fidelity versions. To prevent network effects from affecting our measurements we make sure that the data is present locally at the client before we load it into the application.

We determine the PowerPoint DMI overhead by measuring the time that the PowerPoint export driver spends loading the new slides, *Stage 1*, and cutting and pasting, *Stage 2*, as described in section 3.1.2. We expect that an in-application approach to adaptation would have to perform *Stage 1*, but would dispense of *Stage 2*.

For IE we determine the DMI overhead for upgrading the images in two different ways: *DMI*, which uses DMI to update the images; and *Synthetic*, which approximates an in-application adaptation approach. *Synthetic* measures the time to load and render previously generated pages that already contain the high fidelity images. *Synthetic* is not a perfect imitation of in-application adaptation, because it requires IE to re-load and parse the HTML portion of the page, which an in-application approach could dispense with. We avoid this problem by using only pages where the HTML content is very small (less than 5% of total page size), so that HTML parsing and rendering costs are minimal.

Table 2 shows the results of these experiments. For each policy, it shows the cost of updating a *single* component (*i.e.*, one slide or one image) and the *additional* cost incurred by every extra component that is updated simultaneously. For PowerPoint, the table shows the time in *Stage 1* and *Stage 2*. For IE, the table shows the times for the *DMI* and *Synthetic* implementations.

The PowerPoint results show that the time spent cutting and pasting, *Stage 2*, is small compared to the time spent loading slides, *Stage 1*, which an in-application also has to do. Moreover, the time spent updating the application *Stage 1* + *Stage 2* is small compared to the network time. For example, the average network time to load a slide over the 384 Kb/sec network is 2994 milliseconds, with a standard deviation of 3943 milliseconds, while the average time for updating the application with a single slide is 994 milliseconds, with a standard deviation of 819 milliseconds.

The IE results show that the *DMI* implementation comes within 10% of *Synthetic*. Moreover, image update times are small compared to average network time. For instance, the average time to load an image over 56 Kb/sec is 565 milli-seconds with a standard deviation of 635 milli-seconds, compared to updating the application which takes on average 33 milliseconds with a standard deviation of 19 milliseconds.

The above results suggest that the adaptation cost of DMI calls is small, and that most of the time that it takes to add or upgrade a component is spent transferring the data over the network and loading it into the application, two factors that are expected to be similar whether we implement adaptation outside or within the application.

6.3 Some adaptation policies

We conclude this section by presenting the results, as the end user would perceive them, of some of the Puppeteer adaptation policies we have implemented so far (see Section 3.3). These results also provide some indication of the circumstances under which these adaptations are profitable.

6.3.1 Powerpoint: First slide and Prefetch all text

In this experiment we measure the latency for a Powerpoint adaptation policy that loads only the first slide of a PowerPoint presentation before it returns control to the user, and afterwards loads the remaining slides in the background. We also present results for an adaptation policy that, in addition, fetches all of the text, in the Powerpoint document before returning control. With these adaptations, user-perceived latency is much reduced compared to the application policy of loading the entire document before returning control to the user.

The results of these experiment appear, under the labels *PPT.slide* and *PPT.prefetch*, respectively, in Figures 6, 7, and 8 for 384 Kb/sec, 1.6 Mb/sec, and 10 Mb/sec net-

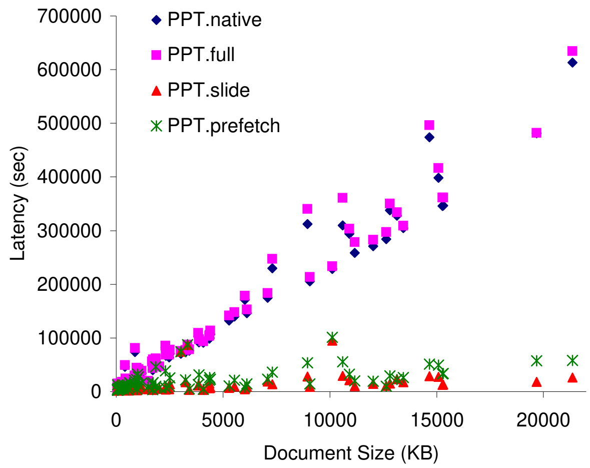

Figure 6: Load latency for PowerPoint documents at 384 Kb/sec. Shown are latencies for native PowerPoint (*PPT.native*), and Puppeteer runs for loading all components (*PPT.full*), loading just the components of the first slide (*PPT.slide*), and loading the text of all slides in addition to all the components of the first slide (*PPT.prefetch*).

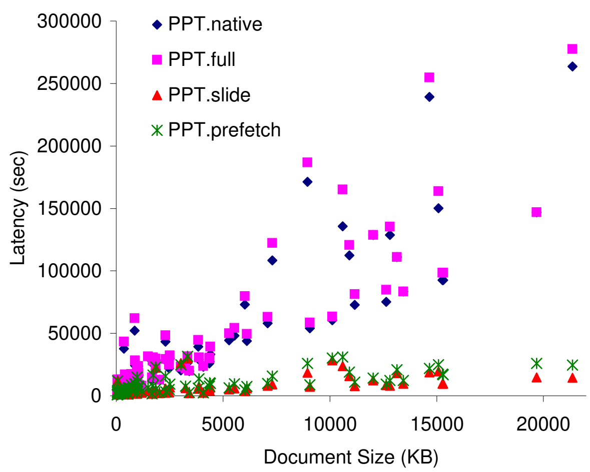

Figure 7: Load latency for PowerPoint documents at 1.6 Mb/sec.

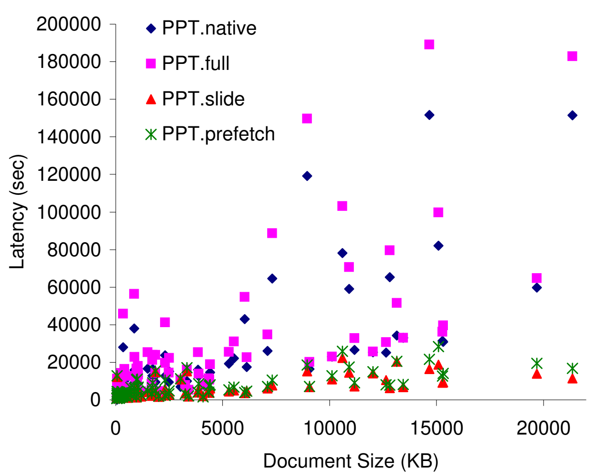

Figure 8: Load latency for PowerPoint documents at 10 Mb/sec.

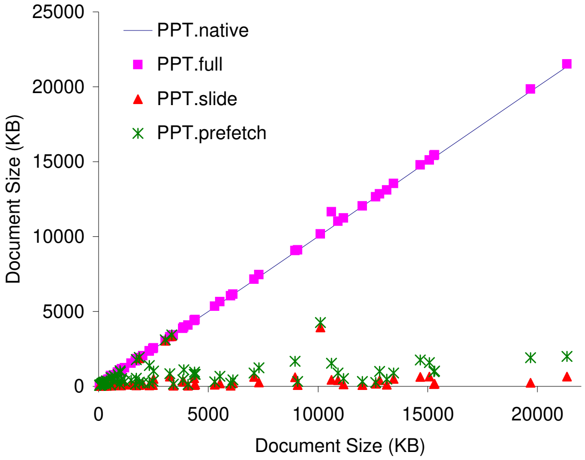

Figure 9: Data transfered to load PowerPoint documents.

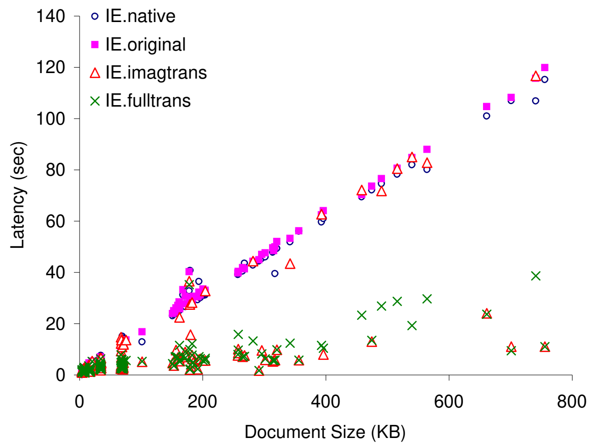

Figure 10: Load latency for HTML documents at 56 Kb/sec. Shown are latencies for native IE (*IE.native*), and Puppeteer runs that load all images at original fidelity (*IE.original*), load only the first $1/7$ bytes of transcoded images (*IE.imagtrans*), load transcoded images and text (*IE.fulltrans*).

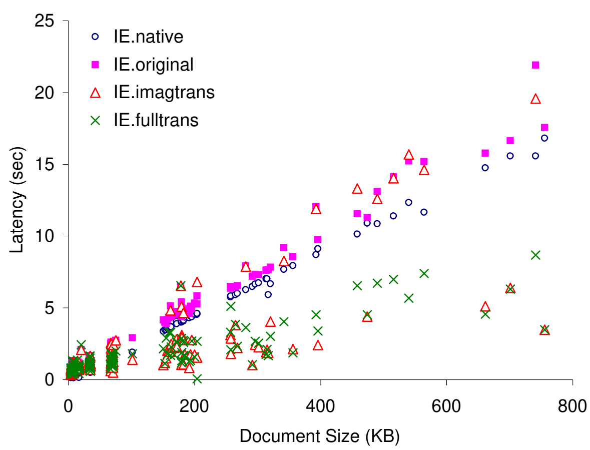

Figure 11: Load latency for HTML documents at 384 Kb/sec.

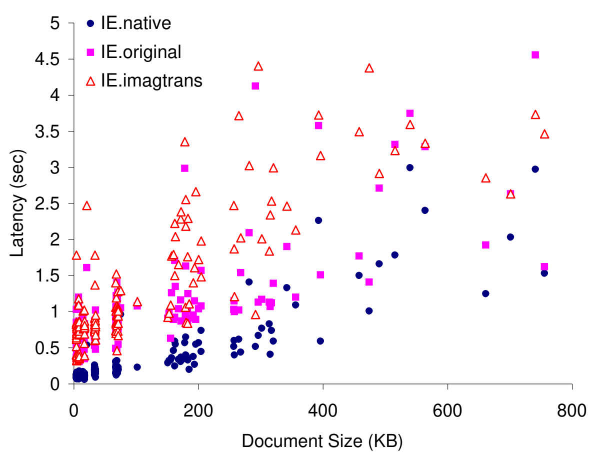

Figure 12: Load latency for HTML documents at 10 Mb/sec.

work links. For each document, the figures contain four vertically aligned points representing the latency or data measurements in four system configuration: native PowerPoint (*PPT.native*), and Puppeteer runs that load all the components of the presentation (*PPT.full*), only the components of the first slide (*PPT.slide*), and the first slide and the text for all remaining slides (*PPT.prefetch*). We chose these particular bandwidths because they capture the transition from the application being network-bound to being CPU-bound. Figure 9 shows the data transfered in each of the four scenarios.

While we expected that reduced network traffic would improve latency with the slower 384 Kb/sec and 1.6 Mb/sec networks, the savings over the 10 Mb/sec network came as a surprise. While Puppeteer achieves most of its savings in the 384 Kb/sec and 1.6 Mb/sec networks by reducing network traffic, the transmission times over the 10 Mb/sec are too small to account for the savings. The savings result, instead, from reducing the initial parsing/rendering time. On average, *PPT.slide* achieves latency reductions of 86%, 78%, and 62% for documents larger than 1 MB on 384 Kb/sec, 1.6 Mb/sec, and 10 Mb/sec, respectively. The data in Figure 9 also shows that, for large documents, it is possible to return control to the user after loading just a small fraction of the total document's data (about 4.5% for documents larger than 3 MB).

When comparing the data points of the *PPT.prefetch* run to *PPT.slide*, we see that the latency has moved up only slightly. The latency is still significantly lower than for *PPT.native*, achieving savings in averages of 75%, 72%, and 54% for documents larger than 1 MB over 384 Kb/sec, 1.6 Mb/sec, and 10 Mb/sec, respectively. Moreover, the increase in the amount of data transfered, especially for documents larger than 4 MB, is small, amounting to only an extra 6.4% above original document

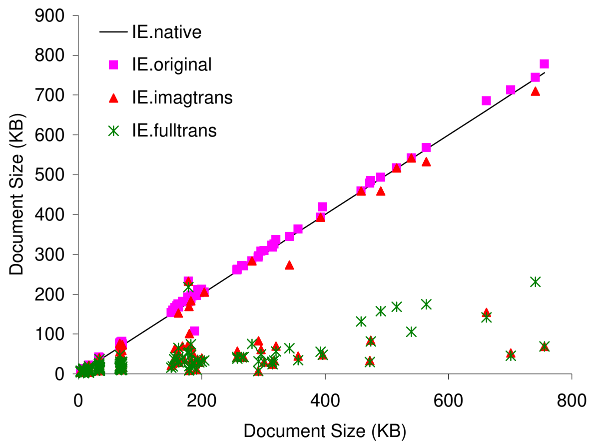

Figure 13: Data transfered to load HTML documents.

size. These results are consistent with our earlier findings of [10] where text accounted for a small fraction of the total data in large PowerPoint documents. These results suggest that text should be prefetched in almost all situations and that the lazy fetching of components is more appropriate for the larger image and OLE embedded objects that appear in the documents.

Finally, an interesting characteristic of the figures is the large variation in user-perceived latency at high network speeds and the alignment of data points into lines as the network speed decreases. The high variability over high network speeds results from the experiment being CPU-bound. Under these conditions, user-perceived latency is mostly dependent on the time that it takes PowerPoint to parse and render the presentation. For PowerPoint, this time is not only dependent on the size of the presentation, but is also a function of the number of components (such as slides, images, or embedded object) in the presentation.

6.3.2 IE: JPEG compression

In this experiment we explore the use of lossy JPEG compression and progressive JPEG technology to reduce user perceived latency for HTML pages. Our goal is to reduce the time required to render a full version of the page by lowering the fidelity of some of the page's elements.

Our prototype converts, at run time, GIF and JPEG images embedded in the HTML document into progressive JPEG format[1] using the PBMPlus [27] and Independent JPEG Group [2] libraries. We then transfer only the first $1/7$th of the resulting image's bytes. In the client we convert the low fidelity progressive JPEG back into normal JPEG format and supply it to the browser as though it comprised the image at its highest fidelity. Finally, the prototype only transcodes images that are greater than a user specified size threshold. The results reported in this paper reflect a threshold size of 8 KB, below which it becomes cheaper to simply transmit an image rather than run the transcoder.

Figures 10, 11, and 12 show the latency for loading the HTML documents over 56 Kb/sec, 384 Kb/sec, and 10 Mb/sec network links. We have again chosen these bandwidths to illustrate the transition from the application being network-bound to CPU-bound. Figure 13 shows the data transfered to load the documents. The figures show latencies for native IE (*IE.native*), and for Puppeteer runs that load all the images at their original fidelity (*IE.original*), load only the first $1/7$ bytes of transcoded images (*IE.imagtrans*), and load transcoded images and gzip-compressed text (*IE.fulltrans*).

IE.imagtrans shows that on 10 Mb/sec networks, transcoding is always detrimental to performance. In contrast, on 56 KB/sec and 384 KB/sec networks, Puppeteer achieves average reduction in latency for documents larger than 128 KB of 59% and 35% for 56 KB/sec and 384 KB/sec, respectively. A closer examination revealed that roughly $2/3$ of the documents see some latency reduction. The remaining $1/3$ of the documents, those seeing little improvement from transcoding, are composed mostly of formated HTML text and have little or no image content. To reduce the latency of these documents we added gzip text compression to the prototype. The *IE.fulltrans* run shows that with image and text transcoding, Puppeteer achieves average reductions in latency for all documents larger than 128 KB, at 56 KB/sec and 384 KB/sec, of 76% and 50%, respectively.

Overall transcoding time took between 11.5% to less than 1% of execution time. Moreover, since Puppeteer overlaps image transcoding with data transmission, the overall effect on execution time diminishes as network speed decreases.

As with PowerPoint, we notice in the figures for IE that for low bandwidths the data points tend to fall in a straight line, while for higher bandwidths the data points become more dispersed. The reason is the same as for Powerpoint: At high bandwidths the experiment becomes CPU-bound and governed by the time it takes IE to parse and render the page. For IE, parsing and rendering time depends on the content types in the HTML document.

6.3.3 IE: Incremental rendering

In this section we extend the *IE.fulltrans* method explained in Section 6.3.2 by incrementally increasing the fidelity of the images displayed by the browser. This experiment comprises a significant departure from the original browser policy that loads single images incrementally,

[1] A useful property of a progressive image format, such as progressive JPEG, is that any prefix of the file for an image results in a complete, albeit lower quality, rendering of the image. As the prefix increases in length and approaches the full image file, the image quality approaches its maximum.

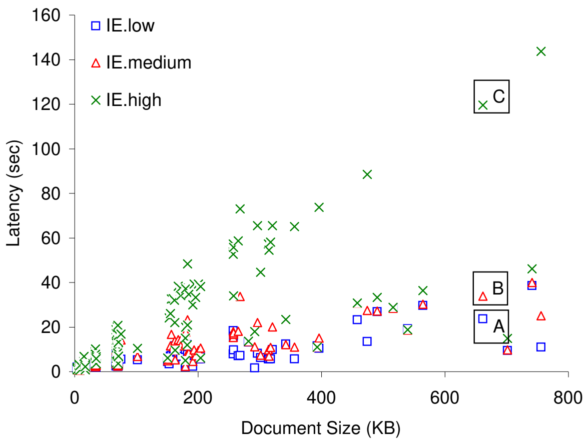

Figure 14: Incremental rendering of HTML documents at 56Kb/sec.

but fetches images in order, with at most four images being loaded at one time. Instead, Puppeteer loads all images at their lowest fidelity first and only then begins to increase image fidelity level.

In this experiment we assume that images have 3 fidelity levels, which we will refer to as *IE.low*, *IE.medium*, and *IE.high*. We chose these levels to correspond to the first $1/7$, next $2/7$, and last $4/7$ of the bytes in the progressive JPEG representation. For GIF images, we first convert them to JPEG. *IE.low* and *IE.medium* are then the same as above. For *IE.high*, however, we transmit the original GIF image. This will support GIF features not present in JPEG, such as animations and transparency.

Figure 14 shows the results of this experiment for loading the HTML document at 56 Kb/sec . For every document, it displays three aligned points with the latency required to load the document at its lowest fidelity (*e.g.*, point A in the plot) and then refined twice (*e.g.*, point B and C). The median time to load all images in a document for documents larger than 128 KB at *IE.low* fidelity was 7.22 seconds. Increasing the fidelity to *IE.medium* and *IE.high* required and addition median time of 3.68 and 22.48 seconds, respectively.

7 Related work

Much work has gone into extending the system architecture to better support mobile clients [20] and to creating programming models that incorporate adaptation into the design of the application [16]. The project that most closely relates to Puppeteer is Odyssey [26], which splits the responsibility for adaptation between the application and the system. Puppeteer takes a similar approach, pushing common adaptation tasks into the system infrastructure and leaving the application-specific aspect of adaptation to application drivers. The main difference between the two systems lays in Puppeteer's use of existing run-time interfaces to adapt existing applications, whereas Odyssey requires applications to be modified to work with it.

Visual Proxies [32], an offspring of Odyssey, implements application specific adaptation policies without modifying the application using interposition between the X-server and the application. While this technique enables many adaptations that are possible with Puppeteer, it requires much more complicated application drivers.

Another project that uses similar ideas to Puppeteer is Dynamic Documents [18]. This instrumentation of the Mosaic Web browser uses Tcl scripts to set the policies for individual HTML documents. While Puppeteer uses the external interfaces provided by the application, Dynamic Documents use an internal script interpreter in the browser.

8 Conclusions

We presented the design and measured the effectiveness of Puppeteer, a system for adapting component-based applications in mobile environments. Puppeteer implements adaptation by using the exposed APIs of component-based applications, enabling application-specific adaptation policies *without* requiring modifications to the application.

We described the architecture of Puppeteer and its implementation. The architecture allows for the modular addition of new applications, component types, transcoders, and policies. We demonstrated that complex policies, that traditionally require significant application modifications, can be implemented easily and efficiently in Puppeteer.

Puppeteer's reliance on application specific drivers to provide tailored adaptation raises the question of porting new application to the system. While we found the effort required to be modest, we are developing a set of standard APIs suitable for adaptation, including facilities for data manipulation and event registration.

References

[1] *GNOME*. `http://www.gnome.org`.
[2] Independent JPEG Group. http://www.ijg.org/.
[3] *KDE*. `http://www.kde.org`.
[4] *KOffice*. `http://koffice.kde.org`.
[5] *StarOffice*. `http://www.stardivision.com`.
[6] BAGRODIA, R., CHU, W. W., KLEINROCK, L., AND POPEK, G. Vision, issues, and architecture for nomadic computing. *IEEE Personal Communications 2*, 6 (Dec. 1995), 14–27.

[7] BROCKSCHMIDT, K. *Inside OLE*. Microsoft Press, 1995.

[8] CHAPPELL, D. *Understanding ActiveX and OLE*. Microsoft Press, 1996.

[9] CUNHA, C. R., BESTAVROS, A., AND CROVELLA, M. E. Characteristics of WWW client-based traces. Tech. Rep. TR-95-010, Boston University, Apr. 1995.

[10] DE LARA, E., WALLACH, D. S., AND ZWAENEPOEL, W. Opportunities for bandwidth adaptation in Microsoft Office documents. In *Proceedings of the Fourth USENIX Windows Symposium* (Seattle, Washington, Aug. 2000).

[11] DUCHAMP, D. Issues in wireless mobile computing. In *Proceedings of Third Workshop on Workstation Operating Systems* (Key Biscayne, Florida, Apr. 1992), pp. 1–7.

[12] FORMAN, G. H., AND ZAHORJAN, J. The challenges of mobile computing. *IEEE Computer* (Apr. 1994), 38–47.

[13] FOX, A., GRIBBLE, S. D., BREWER, E. A., AND AMIR, E. Adapting to network and client variability via on-demand dynamic distillation. *Sigplan Notices 31*, 9 (Sept. 1996), 160–170.

[14] FOX, A., GRIBBLE, S. D., CHAWATHE, Y., AND BREWER, E. A. Adapting to network and client variation using infrastructural proxies: Lessons and perspectives. *IEEE Personal Communications 5*, 4 (Aug. 1998), 10–19.

[15] GARDNER, D. Beginner's guide to DHTML. http://wsabstract.com/howto/dhtmlguide.shtml.

[16] JOSEPH, A. D., DELESPINASSE, A. F., TAUBER, J. A., GIFFORD, D. K., AND KAASHOEK, M. F. Rover: a toolkit for mobile information access. In *Proceedings of the 15th ACM Symposium on Operating Systems Principles (SOSP '95)* (Copper Mountain Resort, Colorado, Dec. 1995), pp. 156–171.

[17] JOSEPH, A. D., TAUBER, J. A., AND KAASHOEK, M. F. Building reliable mobile-aware applications using the Rover toolkit. In *Proceedings of the 2nd ACM International Conference on Mobile Computing and Networking (MobiCom '96)* (Rye, New York, Nov. 1996).

[18] KAASHOEK, M. F., PINCKNEY, T., AND TAUBER, J. A. Dynamic documents: mobile wireless access to the WWW. In *Proceedings of the Workshop on Mobile Computing Systems and Applications (WMCSA '94)* (Santa Cruz, California, Dec. 1994), IEEE Computer Society, pp. 179–184.

[19] KATZ, R. H. Adaptation and mobility in wireless information systems. *IEEE Personal Communications 1*, 1 (1994), 6–17.

[20] KISTLER, J. J., AND SATYANARAYANAN, M. Disconnected operation in the Coda file system. *ACM Transactions on Computer Systems 10*, 1 (Feb. 1992), 3–25.

[21] MICROSOFT CORPORATION. *Microsoft Office 97 Drawing File Format*. Redmond, Washington, 1997. MSDN Online, `http://msdn.microsoft.com`.

[22] MICROSOFT CORPORATION. *Microsoft PowerPoint File Format*. Redmond, Washington, 1997. MSDN Online, `http://msdn.microsoft.com`.

[23] MICROSOFT CORPORATION. *Microsoft Office 2000 and HTML*. Redmond, Washington, 1999. MSDN Online, `http://msdn.microsoft.com`.

[24] MICROSOFT PRESS. *Microsoft Office 2000 / Visual Basic Programmer's Guide*, 1999.

[25] MUMMERT, L. B., EBLING, M. R., AND SATYANARAYANAN, M. Exploiting weak connectivity for mobile file access. In *Proceedings of the 15th ACM Symposium on Operating Systems Principles* (Copper Mountain Resort, Colorado, Dec. 1995).

[26] NOBLE, B. D., SATYANARAYANAN, M., NARAYANAN, D., TILTON, J. E., FLINN, J., AND WALKER, K. R. Agile application-aware adaptation for mobility. *Operating Systems Review (ACM) 51*, 5 (Dec. 1997), 276–287.

[27] POSKANZER, J. PBMPLUS. http://www.acme.com/software/pbmplus.

[28] RIZZO, L. DummyNet: a simple approach to the evaluation of network protocols. *ACM Computer Communication Review* (Jan. 1997).

[29] ROBERTS, S. *Programming Microsoft Internet Explorer 5*. Microsoft Press, 1999.

[30] SATYANARAYANAN, M. Hot topics: Mobile computing. *IEEE Computer 26*, 9 (Sept. 1993), 81–82.

[31] SATYANARAYANAN, M. Fundamental challenges in mobile computing. In *Fifteenth ACM Symposium on Principles of Distributed Computing* (Philadelphia, Pennsylvania, May 1996).

[32] SATYANARAYANAN, M., FLINN, J., AND WALKER, K. R. Visual proxy: Exploiting OS customizations without application source code. *Operating Systems Review 33*, 3 (July 1999).