



White Paper

Crawford Technologies Inc.
45 St. Clair Avenue West Suite 102
Toronto, ON, Canada, M4V 1K9
Tel. +1-866-679-0864
www.crawfordtech.com
info@crawfordtech.com

Is PDF an appropriate choice for high volume transactional production printing?

Even as far back as “Acrobat Reader V4.0”, people have been asking that question. With its free viewers that allow you to easily see the contents of a printed page, people have wondered “Could I use PDF to replace the existing AFP corporate print architecture we use today?” In this document we compare PDF to AFP, the prevalent corporate print architecture, to evaluate this question.

The topic continues to be popular, and has led to several very well attended sessions called the “Print Stream Shootout” at Xplor Global Conferences over the past years. As usual, with any kind of question, it rarely results in a simple “yes” or “no”, but the more commonly used “it depends”. If you are making a strategic decision, we recommend that you read this entire document and urge you to consider all of the points raised here before you make that decision.

The Advantages of PDF

PDF, as a page description language, certainly has some advantages:

1. Like AFP, PDF is page independent, which means that you do not need to process pages 1 to 999 in order to process page 1000. Each page stands on its own. This is valuable when it comes to printing performance. If necessary, multiple processors can be employed to process pages in parallel, improving printer RIP times. While not common, this may be necessary to get PDF files printing on 2000+ image per minute printers.
2. The PDF imaging model (the set of commands that allows it to put text, images, graduated screenings, Bezier curves, vector graphics, text along an arbitrary path, even full color and color management) is top notch. The quality of the dots on the printed page is as good as you’re going to get. In fact, the imaging model that PDF provides probably supports more functions than most transactional statements require: When was the last time you saw variable data written along an arbitrary curve on your phone bill?
3. PDF viewers are everywhere (Windows, Mac, Linux, even portable devices), they’re all pretty good, and they’re all pretty much free. Viewers for other print streams including AFP are not as widespread or as consistent;
4. PDF supports compression of fonts and pages inside the PDF file, and even supports sub-setting fonts (pruning of characters from an embedded font to remove characters that are not required for this PDF document) to make the file smaller.
5. If a font is not present in a PDF file, a PDF viewer will do its best to pick a font that is as close as possible, and substitute it. This is a nice feature, however, if the PDF is being printed, the substituted fonts may not be acceptable, and this will only be found by inspection (hopefully before the documents are distributed to recipients).

These features have made PDF popular as a document viewing and exchange platform, but is PDF an appropriate choice for high volume transactional production printing? The fact is that there are some pitfalls that may make PDF less attractive for this purpose.

PDF in a High-Volume Transactional Printing Environment

There are a number of issues which make PDF challenging to use as a foundation for a high volume transactional document strategy. Since it is rare that 100% of the recipients are prepared to accept electronic delivery of their documents, a fairly large portion of the documents are still being printed, and will be in the foreseeable future. This raises a number of issues including the following:

1. PDF has no defined support for simplex or duplex inside the print stream. Although you can probably force a printer that supports printing of PDF files to run in duplex mode, you will have to add blank back pages into the PDF file in order to start on a facing page (if, for example a customer's document is only 3 pages long). These blank backs may be annoying to a person viewing their statement, so you may need to actually keep two copies of the PDF file around. One that is ideal for printing, and one that is meant for viewing. Furthermore, there is no way in the PDF file itself that you can ensure that the document actually prints in duplex mode; you're simply counting on the operator to set the printer up correctly for this job.
2. PDF has no defined support for paper tray pulls. This also leads to a dilemma similar to the above. Usually, you put paper tray pulls in the print stream so that you can choose pre-printed paper stock such as company letterhead. This then implies that the PDF file has something missing from it (color logos, colored paper, security paper such as check stock, boiler plate text) that make the PDF incomplete (it's missing the logos because they're on the paper), and now, the PDF may be inappropriate or incomplete for viewing by itself. This also means that you probably need two versions of the PDF: one for printing (without logos, static text and color) and one for online viewing (with full color logos, static text and color).
3. The performance of a PDF print job is usually an order of magnitude slower than that of print streams that are geared towards transaction oriented output (i.e. AFP). These transaction orientated print streams were created to drive very high speed digital printers with completely variable data on every page. These speeds are currently around 2000+ impressions per minute. An AFP/IPDS printer can capture and keep resources in memory across multiple jobs so that they do not have to be re-downloaded over and over again. Also, by intentionally keeping the graphics objects in AFP (GOCA) very light and simple, it is easy for the printer controller to render them at rated printer speed.
4. PDF print streams offer no print integrity, (otherwise known as the "fire and forget" approach). In contrast, the AFP architecture includes a bi-directional communication (known as IPDS, or Intelligent Printer Data Stream) between the print server and the print device. This means that if the printer has a paper jam, it reports it to the print server, which backs up and reprints the damaged pages. Furthermore, if the application tries to print outside the valid printable area, or if a required font is missing, then an AFP device will notify you of the error, and terminate the job. PDF printers will usually just ignore the error, or substitute another font. Substitution of a font can result in a serious error if something like a signature on a check is presented on the page using a font. Although this is probably good for ad-hoc viewing of a PDF document, you need to decide for your organization if this is appropriate behavior or not for your production statements.
5. PDF support for metadata is somewhat restrictive compared to AFP. In AFP you can place index (TLE) records anywhere in the print file, and have them associated with a page. You can also associate the index records with a range of pages. There are a number of ways that you can group a number of related page ranges to delimit documents such as statements, bills and notices. In PDF, there are annotations such as Bookmarks and XML objects (XMP) that you can use to associate metadata with pages within the PDF. If you are going to use large PDF files for containing multiple documents such as statements, bills, notices and such, you will need to have an external index in order to track the documents within the PDF file, as there are no elements to group page ranges.



White Paper

Crawford Technologies Inc.
45 St. Clair Avenue West Suite 102
Toronto, ON, Canada, M4V 1K9
Tel. +1-866-679-0864
www.crawfordtech.com
info@crawfordtech.com

6. PDF version 1.7 has a 10 GB limit and older versions have more restrictive size limitations. We have customers who regularly process AFP files that are larger than 20GB. The PDF limits may impact how print runs are handled, as it may force large runs to be split up into multiple PDF files.
7. PDF has no defined support for finishing options such as folding, jogging, stapling and stitching.
8. PDF does not have a spot color space or highlight color space. This means that your PDF files will need to be either black and white or full color.
9. Because of the large PDF files that can result when embedding fonts, many PDF creation software vendors allow fonts to be pruned. This means removing characters from the embedded font to only include the characters that are actually used in that PDF file. To complicate this, the code points normally used when font subsetting is done may not be consistent from file to file. That is, they do not use any predictable code points. This means that PDF files using subset fonts usually can not be concatenated back together again (i.e. to print in a large batch), because now each font in each PDF file is completely unique. The resulting PDF could have thousands of fonts in it, and post-processing software used to index the files will not work with them.
10. Adobe has demonstrated a lack of commitment to backward compatibility: Over time, Adobe has changed the Acrobat Reader to cause documents that used to be viewable using Acrobat version <x> to no longer be viewable using Acrobat version <x+1> (Specific example where x=7). This is especially troubling if you plan on using these PDF documents as your long term archive, and documents from 5 years ago no longer view with the current version of Acrobat.

The above list includes some of the most important issues to think about when considering whether or not to use PDF as your production print stream standard for high volume transactional printing. As mentioned earlier, some or all of the issues mentioned above may or may not apply to your specific installation.

The Ongoing Need for PDF in a High Volume Transactional Print Environment

Having said all of the above, it is becoming more and more common for print service providers or corporate and government in-plant print shops to have a need to bring PDF documents into their existing production architecture. Crawford Technologies has recognized this and provides market-leading software tools to convert PDF into other production print architecture formats such as AFP, Xerox Metacode, PostScript, PCL and repurpose them for production printing. This includes being able to look for text/data in a document and/or external file so that the software can insert the appropriate commands to pull that particular page from a specific paper tray, make a particular page the front side or the back side of a page or perform finishing functions in the resulting print file.

If your organization decides that PDF is the way to go, then here is how you can best accomplish it:

1. First, accept that you will likely need at least two versions of each file, one for printing and another for archive/viewing purposes. While this may be counter to the original reason for you to use PDF for printing, you need to accept that the requirements for viewing and printing are quite different, and in order to achieve success without compromise, you will need two versions. You may also need to create a PDF/A version of those PDF files that you are going to archive for the long term. Software such as a CrawfordTech's PRO Transform has the ability to convert a PDF from one of these versions to the other, so you do not need to get the producer to create both versions. It can also add required blank sides to properly print duplex and add barcodes required for document processing facilities and inserters. If you are converting from a PDF optimized for printing, it can remove the



White Paper

Crawford Technologies Inc.
45 St. Clair Avenue West Suite 102
Toronto, ON, Canada, M4V 1K9
Tel. +1-866-679-0864
www.crawfordtech.com
info@crawfordtech.com

unnecessary blank pages and suppress unwanted barcodes for the online versions of the PDFs. It will also convert standard PDF files into PDF/A for archival purposes.

2. When printing large quantities of customer communications documents, it is not reliable or efficient to print single document PDF files, as managing hundreds of thousands of PDF files for a single print run is very troublesome. Thus, you will need to batch together a large number of documents into one PDF file. This batching process must be done in a way that creates an efficient PDF file. Since the entire PDF file must be resident on the printer before it can start printing, there may be some printer memory and network performance implications that need to be considered when setting up your production print facility.
3. Depending on your print volumes and throughput requirements, and the vendor or device type you will be printing on, you may still need the actual printing to be done using a language such as AFP. This may be done using a print stream transform to convert the PDF to AFP. You need to be careful when you put this in place. Many such transforms create full page raster images for the pages. This can result in very large AFP files that use up a large amount of disk space, take a long time to transmit, and may slow down the printer. If you are using CrawfordTech's PRO PDF product, it does an "object to object" translation wherever possible, keeping text as text which minimizes the output file size. There are some other things to consider if you plan on converting your PDF print streams to AFP for production printing:
 - If you are printing on cut sheet printers, and your application uses multiple paper stocks, you will also need something to insert the bin or paper tray pulls, as PDF does not support these. The same is true if you plan on running pages in a mix of simplex and duplex environments in the same print run. Intelligent document processing software like CrawfordTech's PRO Document Enhancer family can provide this functionality.
 - Due to the complex imaging model of PDF, conversion programs can be very slow, so you need to test the performance of the conversion with your applications so that you can be sure you have time in your production window for the conversion to complete without delaying your print schedules.
4. For long term storage and archival of your PDF files, you have three potential strategies to choose from:
 - You can store the files in PDF or PDF/A format. PDF/A is a version of PDF that is optimized for long term storage. It is designed so that they can be read in the future and will always be formatted correctly.
 - You can store individual files for each document (statement, bill, invoice, notice, etc.),
 - You can store large collections of documents in a single file (normally done with a print job).

In general, PDF/A files are larger than PDF files, and storing individual document files are much larger than storing a single file with a collection of documents in it.

Most Enterprise Content Management systems will handle individual PDF and PDF/A document files, and some of them support multi-document PDF and PDF/A files, and have the necessary components to retrieve individual documents when needed.

Conclusion

Upcoming standards such as "PDF/VT" and CIP4's "Common metadata for Document Production Workflows" are trying to address some of the issues described above. Industry adoption and acceptance of these Graphics Arts industry standards, when they are finalized and published, will determine whether



White Paper

Crawford Technologies Inc.
45 St. Clair Avenue West Suite 102
Toronto, ON, Canada, M4V 1K9
Tel. +1-866-679-0864
www.crawfordtech.com
info@crawfordtech.com

these issues will be resolved. It is clear however, that such industry change will take time, and until such time has passed, these issues remain.

In the future, it may be possible to combine PDF and AFP in a unique manner to achieve the best of both worlds. The AFP architecture has been extended to allow an entire PDF file to reside in a single Object Container. Pages within a multi-page PDF in an Object Container can be referenced from pages in the AFP print file. Once the technology gets to the market to create, manage and print these documents, then it may be viable to wrap AFP around a PDF file and achieve high speed printing with the integrity and robustness of AFP and the imaging model of PDF.

In the meantime, PDF files themselves can be used for a transactional customer communications print strategy as long as the right tools are available for the necessary conversions and document re-engineering needed to provide efficient printing and archival of the documents. These tools are available from vendors such as Crawford Technologies Inc. that allow you to re-engineer your print architecture for new generations of applications.

About Crawford Technologies Inc.

Since 1995, Crawford Technologies' award-winning solutions have helped over 700 companies around the world reduce costs associated with communications processes by delivering bills, statements and other mission-critical transactional communications to their customers in the format they need, when they need it. This includes alternate format documents in Braille, Large Print, Audio and E-text for visually impaired and print-disabled customers.

With CrawfordTech's range of unique software products and services, our clients simplify, automate and extend document delivery cost effectively - irrespective of current, legacy or future standards in infrastructure or document output. These clients, including four of the top five US banks, four of the top five US insurance companies and four of the world's top five car manufacturers, are realizing high-value results as costs associated with document processes are minimized through automation, new opportunities for savings across critical communications are realized and they react quickly to changes in regulations, policies, business requirements and technical infrastructure.

CrawfordTech's quality software, expert support and print business knowledge help clients to meet operational, service, marketing, legal and standards requirements and automate the delivery of billions of communications to their customers annually. Simply put, companies look to Crawford Technologies for a platform-independent approach, leading system performance and superior output fidelity. Please visit www.crawfordtech.com to find out more about CrawfordTech's clients, people, partners and solutions.