Embedded Scanning, Encryption and Certification Workflows on Multi-Function Printers (MFPs)

Helen Balinsky and Nassir Mohammad
Hewlett-Packard Laboratories
Long Down Avenue, Bristol BS34 8QZ, UK
E-mail: Helen.Balinsky@hp.com and NMohammad@hp.com

Abstract

A recently proposed document format: Publicly Posted Composite Documents (PPCDs) [1] has been developed to tackle both the containment and security aspects of composite documents participating in inter-organizational workflows. Here, the access control and authentication material together with the actual contents form a tamper proof digital bundle that was designed to be securely transmitted over potentially low security channels. Though PPCDs can be created, shared cross-organizationally and edited using software on a computer, a key functionality to efficiently and securely incorporate scanned documents inside Multi-Function Printers (MFPs) was missing. We solve this problem by enabling users with assigned modification rights to specific content-parts of a PPCD, to update such parts through scanning physical documents at a standalone MFP. The scans are then securely and automatically encrypted, signed and inserted as new PPCD content. Thus, in this paper we provide a comprehensive solution for document workflow participants to securely update or insert new contents to the PPCD at a standalone MFP device through non-trivial extension of HP MFP firmware.

Introduction

Document workflows are an integral part of many types of organizations including healthcare, governments, large corporations and small and medium businesses. These workflows are not limited to secure intra-organizational scenarios, but are rather often cross-organizational where information is required to flow both in and out of an organization over insecure channels. Furthermore, workflows often round trip many times between the digital and physical realms through scanning, digitization and printing. Although multi-function printers (MFPs) are an integral part of the physical to digital document crossover workflows, they currently offer little for assembly and security of multiple scanned documents.

One can consider many sensitive workflows (e.g. mortgage applications, patent forms, authorization processes) that require individuals to print, fill in and sign forms, and at the same time provide authentic scans of supporting or valuable originals (e.g. passports, receipts, utility bills, pay-slips). Thus, a workflow step may involve several and different types of documents to be created or sent together, and this can lead to problems in keeping the required documents together without any chance of separation.

Composite document formats can resolve this issue to some extent, however, when security is also of concern, the document contents may be required to be only accessible to a select few individuals, groups or locations. Furthermore, any changes that are made to such composite documents by an authorized person should be verifiable by another workflow participant. The absence of a remote authority, particularly between different organizations makes the security requirements difficult, if not impossible, to uphold.

In order to address the concerns of cross-organizational sharing and access of documents, Publicly Posted Composite Documents (PPCDs) [1] is a technology that was introduced to provide the containment of potentially sensitive documents together with the access rights as an integral part of the document. This digital bundle is also signed, with signatures verifiable by all authorized agents and does not require any remote server control and overhead to administer access rights. Furthermore, due to the PPCD being an encrypted format it can traverse between different organizations, multiple devices and regions without demanding any secure communication channels, and hence the compromising of document contents.

However, though PPCDs can be utilized on a computer using appropriate software, the addition of physical versions of a document requires scanning on a MFP. This raises an unacceptable security issue of potentially sensitive document contents being sent in unencrypted form and over insecure channels between the MFP and the computer hosting the PPCD. The absence of a secure and differentially accessible format inside a device means that scans of important documents cannot be securely added to a document, except with the contents being encrypted by the device and sent to the user's computer where they can be decrypted and added to the PPCD. However, this still potentially can leak the data as once decrypted, the sensitive document may be unacceptably in plaintext form on the user's computer. Additionally, coupling of a computer with a MFP complicates matters since ideally the process should be streamlined so that the MFP performs all the necessary processes.

Existing Solutions

To the best of our knowledge, there currently are no existing solutions that enable MFPs to encrypt, sign and incorporate document scans to a PPCD inside the device. However, there are solutions that aim at streamlining, automating or securing document scanning at a MFP. Of the existing solutions, HP Capture and Route [2] provides a fast and easy way to scan, process and route documents, but does not provide a way to handle the security aspects of sensitive documents at the device. Xerox Scan to Desktop [3] provides users a secure way to scan documents to Microsoft Office and PDF formats for fast processing and routing. However, the available formats do not cater for certification or composite document workflows. In the case of certifying documents, scans of valuable originals may be combined and certified into a bundle with other office documents at a MFP device through emails sent with a MIME signature. However, note that such signatures may not always be kept together with the composite document as they do not form an integral non-separable part.

The current scanning solutions aim at solving aspects of workflow problems but lack a unified secure framework for simultaneously originating and editing differentially accessible composite documents, providing certification at the point of scan without reference to an external service, and ensuring document contents cannot be viewed by non-authorized agents.

Contributions of the Paper

The goal of this paper is to show how MFP device firmware can be extended to utilize the recent PPCD technology for executing secure and streamlined document scanning workflows. We turn an MFP into a 1st class device that can originate PPCD workflows using scans of physical documents, and which can help users process scanning workflows on documents that arrive at the device. We let PPCDs serve as a secure scan serialization inside a device for differently formatted scans to be combined into one file. Recalling that PPCD content-part access for each user are of three types: v*erify/only, read/only,* and *read/write*, the *read/write* parts are the only modifiable parts that don't cause the breaking of the secure workflow, and hence the only parts that trusted workflow participants can securely scan to or update from a list of such content-parts.

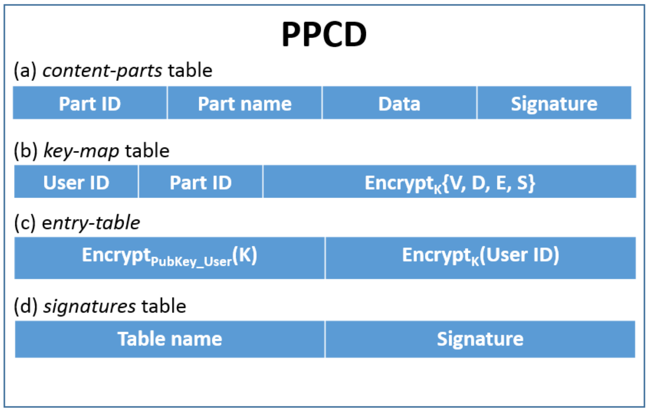
Our complete solution, enables MFP devices to validate and authenticate PPCD's, provides users the ability to scan particular documents and automatically integrate scans into a composite format to be possibly routed to a specified destination(s) as a single integral unit with built in security. This improves over existing solutions since it provides a secure framework for document authentication on arrival by a device, differential access, user prompts, scanning, encryption and certification of documents – all at a single standalone device. Furthermore, PPCDs can now originate at a device from physical versions, with the content-parts certified and encrypted internally at the point of scan – without reference to any external devices or services. The security focused and differentially accessible format makes this a challenging and non-trivial task to implement inside a MFP device that is typically power constrained by business cost requirements.

Paper Outline

The rest of the paper is as follows: for user convenience we start by recapitulating the key components of the PPCD format and, in particular, authorized content modification whilst preserving document authenticity. Then we outline the key problems to solve and present our MFP scanning, encryption and certification solution architecture followed by implementation details of the novel aspects. This is followed by a discussion of our conclusions and avenues of future work.

PPCD Brief Overview

PPCD technology has been described in the paper [1] in great detail. Hence, in this section we briefly recall the key components of PPCD structure for user convenience, and illustrate the modifiable parts of the document. A PPCD is a serialization over a SQLite flat file database containing natively formatted documents (e.g. Microsoft Word, Adobe PDF, text file, etc.) that may be encrypted and signed. Due to its SQLite nature, PPCD serialization is composed of several tables that correspond to the accessibility, differential access control of content, and the content itself. The relevant fields of these tables are illustrated in Fig. 1.



PPCD

(a) *content-parts* table

Part ID	Part name	Data	Signature

(b) *key-map* table

User ID	Part ID	$\text{Encrypt}_K\{V, D, E, S\}$

(c) *entry-table*

$\text{Encrypt}_{\text{PubKey\_User}}(K)$	$\text{Encrypt}_K(\text{User ID})$

(d) *signatures* table

Table name	Signature

Figure 1. PPCD tables overview of relevant fields. A PPCD is composed of several tables with fields that hold the document contents and the access meta-data. This table illustrates the tables and main fields which compose the document format.

The *parts* table (Fig. 1.a) holds the content-part ID, content-part name, possibly encrypted and signed content-part data, together with their signatures and any other optional associated fields, e.g. messages to be displayed when the content-part is selected, timestamps, notes, etc. Each content-part in a simple workflow can be assigned 4 unique access keys: encryption key E, decryption key D, signature generation key S and verification key V. Each workflow participant with access to a PPCD is given a subset of content-part keys corresponding to the access granted by the document creator, and for each workflow step. Every workflow participant is given the verification key V which allows them to only *verify* the validity of the corresponding content-part. Participants granted *read/only* rights are given the subset {V, D} which allows them to additionally decrypt and view the plaintext content. A participant that is granted full access, i.e. *read/write*, is provided the full set of keys. Note that it is only users with *read/write* access to a particular content-part that are able to perform modification without breaking document authenticity.

A *key-map* table (Fig. 1.b) is used to hold the subset of keys for every PPCD content-part and provides for the user's document access. Each workflow participant is identified in the rows of the *key-map* table by their User ID, and thus have their own unique access rights to document contents. Note that the *key-map* table is an integral part of the PPCD with the access keys for each content-part encrypted using the participant's own unique and corresponding symmetric key K. This ensures that only the possession of a key K allows recovery of the participant's document content-part keys, and hence of the plaintext content.

The keys, K, used to encrypt a participant's access keys in the *key-map* table rows, are stored in a row of a filtration table called the *entry-table* (Fig. 1.c). This table contains rows corresponding to each workflow participant, whose *key-map* table User ID is encrypted by key K. Each row also contains the key K which is itself encrypted using the workflow participant's public asymmetric key. Thus, only participants with a corresponding private key can decrypt and recover their key K.

The *key-map* and *entry-table* provide access control of PPCD content-parts. However, in order to ensure document validation can be performed by every workflow participant, signing and verification processes are used in PPCD creation/access protocol. The *key-map* table and *entry-table* in the document are signed by the document master using their private key. These signatures are stored in a separate *signatures* table (Fig. 1.d) that is itself also signed by the master. The corresponding public key certificate of the document master may then be embedded within the document itself, and verifiable through a certification chain or using a common certificate authority (CA) that has signed the document master's certificate. Furthermore, in order to ensure that every PPCD content-part cannot be modified by unauthorized agents, each part is signed using a signature key whose corresponding verification key is provided in each participant's row of the *key-map* table. In deployment scenarios the process of verification on PPCD access ensures that a workflow participant can be satisfied that the PPCD they are using is an authentic version.

The Problem

Several challenges need to be addressed in order to enable workflow participants to securely modify PPCDs using document scans at MFP devices. The problems are further exacerbated by limited resources and power of MFP devices due to business cost constraints. The security-driven nature of the problem also means that every aspect of our solution must ensure a secure system that does not compromise sensitive documents, user identity credentials or any cryptographic materials (e.g. keys).

Our solution must ensure that MFP devices are enabled to accept PPCD-formatted jobs without automatically submitting them to the printing pipeline: the actual job contents is sensitive and not accessible without an authorized workflow participant. Our solution should not interfere with the printing or processing of other jobs arriving at the device, such that if a non-PPCD job is encountered, the data should be returned back to the input stream unaltered, to be handled by its dedicated solution.

Once documents are stored on the device's hard drive, authorized users must be provisioned with a way to navigate the device's control panel UI to select a method for engaging their private key or on-line identity that enables document access. The UI must also enable users to navigate to stored documents, select a document, and select specific content-parts of a PPCD-document to add scans to. Furthermore, the UI should also provide important information to the user so that document scanning can be performed correctly. This may include providing a list of accessible PPCDs, lists of content-parts permitted for update within a selected PPCD, as well as information about the content-parts themselves with granted access. The UI should also provide information about the scan options and processing, together with messages informing about scanning or workflow completion.

Verification of PPCD authenticity, access by workflow participants, access to modifiable content-parts, and subsequent replacement of content-parts by encrypted and signed scans into the PPCD serialization needs to all be implemented in the device's firmware. It may not be outsourced to an external server/service to ensure that sensitive contents never leave the secure and trusted device. On the other hand, a fully embedded solution without any external server or service is easier to maintain, and less vulnerable due to the reduced surface of attacks. Furthermore, in many countries, especially in Europe, outsourcing of the encryption, and/or signature from the MFP may invalidate workflow compliance.

The management of sensitive data in the device is an important consideration that must also be addressed. Any private keys or passwords brought or entered at the device must be handled securely. Furthermore, any subsequent keys that are decrypted for accessing the document and its content-parts must be managed appropriately, i.e. accessed only when required, securely handled, and discarded immediately after being used. Scans of physical documents must also be handled securely and timely discarded to prevent any leakage of potentially sensitive data.

Scan-to-PPCD Solution Overview

Scan-to-PPCD workflows are enabled on the MFP device through a number of firmware extension modules that fit together to provide a comprehensive solution from receiving PPCD formatted jobs, through to workflow completion. Scanning-to-PPCD solution is comprised from the following key components (also shown in Fig 2):

Dedicated PPCD Input/output Filter module (Fig. 2.c), which identifies and processes PPCD-formatted jobs only (Fig. 2.b), whilst other jobs are returned back to the input data stream for processing by their assigned filters (Fig. 2.d). PPCD jobs can be retrieved or sent to a device from a variety of sources such as laptops, shared drives, cloud services or USB drives (Fig. 2.a). The filter stores PPCD-formatted jobs in the dedicated storage area in the device's hard drive (Fig. 2.e).

Device's Control Panel navigation and UI module provides users with the navigation control (Fig. 2.k) where a method for private key retrieval/access can be selected (Fig. 2.g), and stored PPCD jobs (with granted access) can be selected for updating. This module is also responsible for providing a user with informative displays upon content-part selection. Once workflows are completed users may also be shown document export options.

User private key access module handles the retrieval/access to the user's private key from an online storage or portable memory medium. It may also be engaged with the device using a smart card reader (Fig. 2.f). The user maybe prompted through the CP for access credentials.

Authentication and Access module checks document authenticity following standard PPCD authenticity protocol [1]. This also simultaneously checks that the user has access to the selected document using their private key (Fig. 2.h).

Scanning module initiates and executes the scanning of physical documents (Fig. 2.l), which are temporarily stored (Fig. 2.m) in a dedicated location on the device's hard drive.

Content-part insertion module encrypts and signs a scanned document by the corresponding key assigned to the selected content-part. The encrypted part and its signature are inserted into the PPCD serialization (Fig. 2.i).

Workflow Completion Module informs users via the CP about workflow completion, and ships the updated PPCD document out of the MFP device by the user identified route (e.g. e-mailing, uploading to a share drive, etc.) The module subsequently ensures that all sensitive data is securely discarded after the workflow participant completes their step or when any error is encountered by the device during document processing.

Our solution modules work together to provide the necessary functionality for completing scanning workflows. The following steps detail the procedure, whilst Fig. 3 illustrates the key workflow components from the user's point of view.

Step 1. Receive a PPCD document and engage user's access key

Step 2. Prompt selection of a PPCD document at the device's CP.

Step 3. Verify authenticity of PPCD tables, if failed: inform the user and stop.

Step 4. Retrieve user's symmetric key K from the document *entry- table.*

Step 5. Decrypt, using key K, to recover the access keys to each content-part so that the verification keys can be used to verify authenticity of each content-part. If checks fail, inform the user and stop.

Step 6. Using user's secret key K identify content-parts with *read/write* access granted and present the list of content-parts by name to the user, providing them with options to select one or a few for part replacement.

Step 7. Show a preview of newly scanned contents on the CP and ask user's confirmation for acceptance. Repeat the scanning until the user accepts the new scan.

Step 8. Using user's key K retrieve access keys (E, S) for the content-part.

Step 9. Using E and S encrypt and sign the new content-part and place it into PPCD serialization, replacing the selected content-part.

Step 10. Securely discard the newly scanned content-part, encryption and signature key for this part.

Step 11. Repeat Step 6-10 for each new document to be scanned to the PPCD.

Step 12. Securely discard user's key, K.

Step 13. Proceed to export the updated PPCD from MFP device.

Implementation Details

The Scan-to-PPCD firmware solution has been prototyped for HP MFPs. The implementation details are given here by expanding upon the modules overview introduced in the previous section. The solution is designed using the HP Open Extensibility Platform for devices (OXPd) and embedded Solution Development Kits (SDKs) [4] available on HP MFPs running Windows CE. The details are specific to HP MFP technology, however, we also utilise the open source library BouncyCastle [5] for performing all cryptography functions. Furthermore, since a PPCD is a SQLite database, we also use the open source SQLite C++ library [6] to read and insert data to the tables.

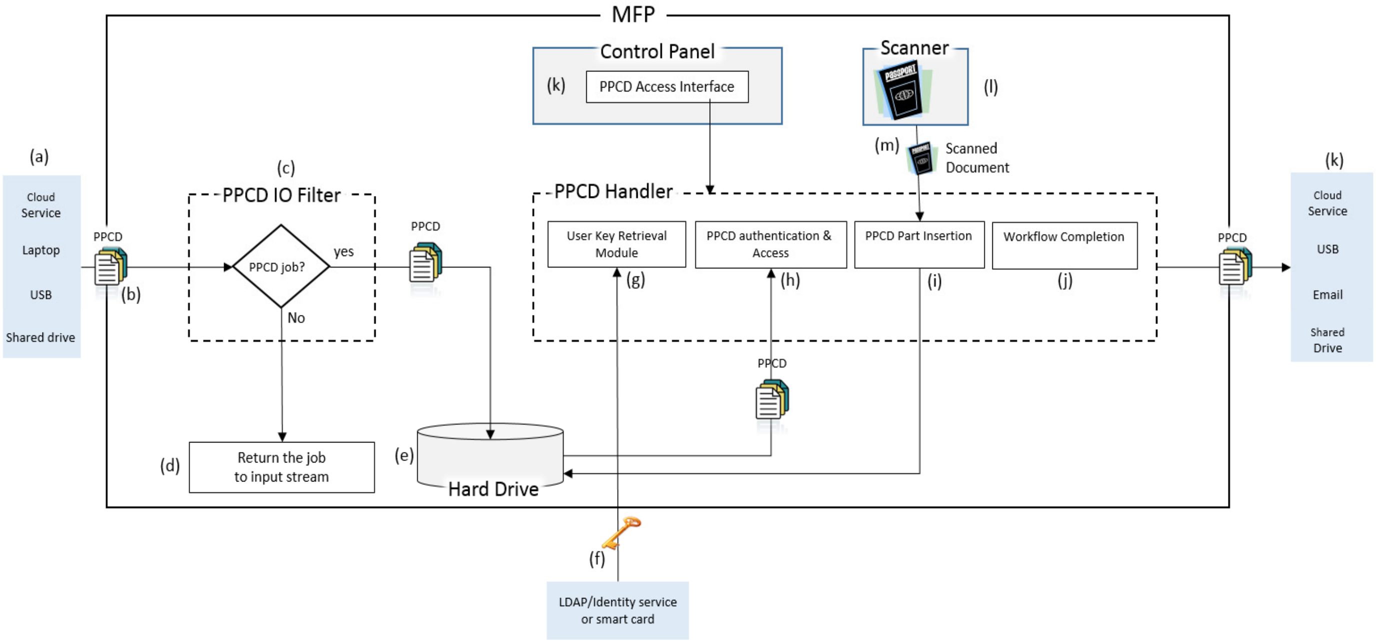


Figure 2. PPCD Scan Workflow. *This figure illustrates the solution architecture inside a MFP device that enables PPCDs to be received, processed, stored, accessed, authenticated, have scans securely inserted and be communicated out of the device upon workflow completion.*

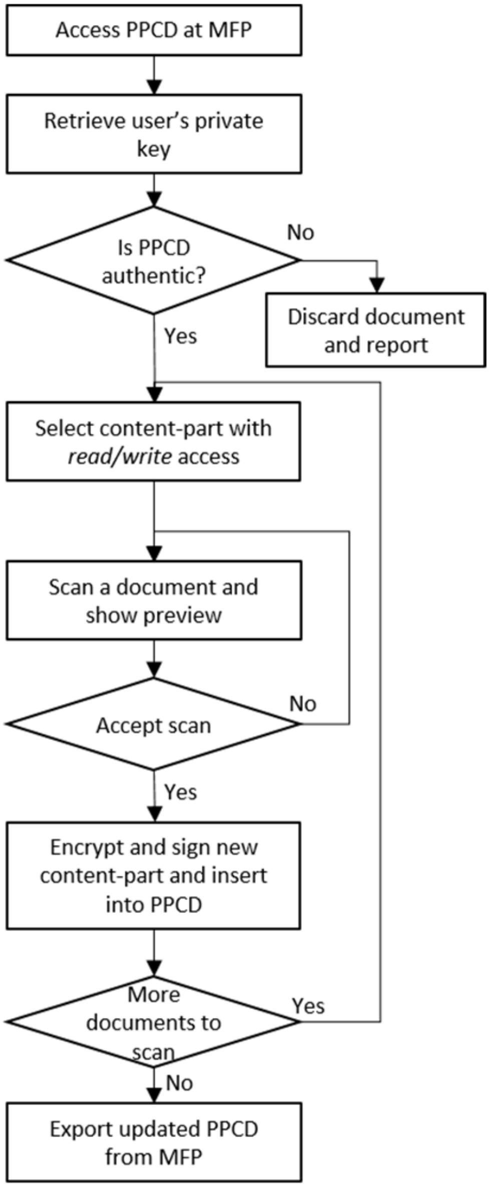

Access PPCD at MFP
Retrieve user's private key
Is PPCD authentic?
No
Discard document and report
Yes
Select content-part with *read/write* access
Scan a document and show preview
Accept scan
No
Yes
Encrypt and sign new content-part and insert into PPCD
More documents to scan
Yes
No
Export updated PPCD from MFP

Figure 3. PPCD Scan Workflow. *This flowchart outlines the Scan-to-PPCD workflow followed by users at a MFP device from document access through to workflow completion and exporting.*

Receiving PPCDs at a MFP

Since PPCD jobs are encrypted, on receiving they cannot be handled directly, but need to be stored by a MFP device until required for access by an authorised user. In order to be identified by the dedicated PPCD-IOFilter the documents are PJL-wrapped containing a unique dedicated command that is only recognized by the PPCD-IOFilter. When other filters encounter this unrecognized command they simply ignore PPCD-formatted jobs, and return them to the input data stream until the PPCD-IOFilter is reached. The dedicated PPCD command contains the job name and size of the PPCD, as well as other relevant metadata. The job name is used as the filename of the PPCD to be stored in a dedicated secure location of the device's hard drive.

Control Panel Navigation and Private Key Retrieval

The CP of the device provides an existing UI that allows users to engage with the MFP. The current prototype provides a user with functionality to retrieve a copy of their private key from a central LDAP repository service using the device's existing functionality. A central key repository is an important use case as it simplifies management of keys in an organization, and provides for timely key revocation by an administration if and when required. It is also the primary mechanism by which the same key can be shared by groups of people, e.g. in the same role, for the same assignment. Security reasons also necessitate private keys to be stored in the LDAP repository and brought to the device in the encrypted PKCS12 format [7]. Users are prompted to provide LDAP repository access credentials (username and password) for their personal or role-assigned private keys to be fetched. In future, devices could also engage the private key via a smartcard reader, so that the private key can be accessed directly from the smartcard. Upon navigating and selecting a document for workflow completion, the device also automatically creates folders in the temporary partition of the device's hard drive for storing sensitive information such as plaintext document scans. The dedicated temporary directory on the device's hard drive is discarded upon workflow completion or error, or when the user signs out of the device.

Document Authentication and Access

Once a workflow participant has selected a document, authentication and decryption processes are carried out. All table signatures and fields (i.e. *entry-table, key-map,* and *signatures* table) are read (using SQLite functions), and verified (through BouncyCastle API) using the document master's public certificate that may be embedded inside the document or available on the device. The certificate may be signed by a trusted CA or otherwise known by a workflow participant. If authenticity of a selected PPCD fails, the user is informed and the document may be discarded. In some cases document authenticity may not be required and the user may be asked to wave this step.

The user's private key is then used to decrypt fields in the *entry-table* according to the PPCD access protocol described in [1] to recover the user's symmetric key K, and the User ID, which is a unique random number assigned to each workflow participant. If the user's key K cannot be recovered, then the user does not have access to the document – the user is notified by the message displayed in the CP. When the user's key K is successfully recovered, it is used to decrypt and recover all the content-part keys (according to keys provided in the *key-map* table, Fig. 1.b). The verification keys are used to verify authenticity of each content-part, while the availability of encryption and signature keys identifies all content-parts that are accessible for modification. Part names (Fig. 1.a) of all content-parts with *read/write* access granted are then displayed on the CP for user selection.

Scanning and Protecting Documents

Our solution utilizes OXPd API for the scanning functionality. When documents are placed on the scan bed and the start button is pressed, this initiates the scanning. Users may be shown previews of the newly scanned content on the device's CP and asked to confirm acceptance of the scan. This may be repeated until the user accepts a scan, and for all content-parts that are to be modified. Once the scan is accepted, the process of the content-part insertion into PPCD-serialization begins according to PPCD update procedure [1]. The corresponding encryption key E and signature key S for the content-part are used to encrypt and sign the newly scanned content. BouncyCastle's AES encryption and RSA/DSA signatures are used in the current prototype. SQLite update commands are then used to replace the encrypted content-part and its newly computed signature in the *parts* table (Fig. 1a).

Workflow Completion and Clean Up

Once the document content-parts have been updated using scans and the workflow completed, the cryptographic keys and any temporary folders that were created to hold scans and encrypted data are securely discarded using OXPd API which specifies the paths and files to be securely deleted. We stress here that any sensitive data brought to the trusted device, or any content that is decrypted and temporarily stored on the device, is in a secure partition that is only accessible through OXPd API.

Workflow completion also provides users with options to export the updated PPCD from the MFP device using the existing digital communication functionality. The user is prompted to select an export path, e.g. upload to SharePoint, cloud service, e-mail or upload to a shared drive. In the current proof-of-concept prototype we only support the emailing of documents which requires prompting the user to enter the destination email address. Once emailed, the user is provided with message displays informing them that the document has been sent. Users may then select more documents for workflow completion, sign out or simply leave the device.

Originating Documents from a MFP

A result of our in device Scan-to-PPCD solution is that PPCDs can now originate at the device. This may be beneficial for initiating some types of document workflows which need to have document scans added by the document master before shipping to subsequent workflow participants. A document master may use their private key and existing PPCD templates (with themselves as the single participant) on device, to create a new document and securely insert scans. Once completed the PPCD may then be emailed to themselves so that the rest of the workflow creation and addition of participants can be carried out on a personal computer as in [1]. Note that where content-parts are to be left empty for workflow participants to insert scans to, they must contain template documents (e.g. passport_template.pdf) which may inform the user about the documents to scan. These content-parts must also be encrypted and signed so that workflow participants can verify the authenticity of such parts.

Conclusion

In this paper we presented a new Scan-to-PPCD solution that can be wholly and securely completed on standalone HP MFP devices using the corresponding firmware extensions. Access to PPCD-formatted jobs on device is restricted to authorized workflow participants only. The newly scanned contents is encrypted and signed/certified at the point of scan inside the MFP, without reference to any external service or device. Scanning workflows are also streamlined with users being able to complete all their actions at MFPs without requiring an additional computer. In future we will look to increase the efficiency of document workflows on MFP devices using PPCDs. In particular, we will study problems in workflow automation and data leak prevention.

References

[1] Balinsky, H.Y. and Simske, S.J. Differential Access for Publicly-Posted Composite Documents with Multiple Workflow Participants, ACM Symposium on Document Engineering, pp.10, 2010, Manchester, UK

[2] HP Capture and Route Solution, http://www8.hp.com/us/en/business-partners/printing-global-partners/hp-capture-and-printing-solutions.html

[3] Xerox Scan to Desktop, http://www.xerox.co.uk/office/software-solutions/xerox-scan-to-pc-desktop/engb.html

[4] OXPd and Embedded SDK overview of proprietary software, http://h71028.www7.hp.com/enterprise/downloads/OXP_brief_4AA0-4557ENW_final.pdf

[5] Bouncy Castle Library, https://www.bouncycastle.org/

[6] SQLite Library, http://www.sqlite.org/

[7] PKCS12 format standard, http://www.emc.com/emc-plus/rsa-labs/pkcs/files/h11301-wp-pkcs-12v1-1-personal-information-exchange-syntax.pdf

Author Biographies

Dr. Nassir Mohammad is currently a Research Scientist in the Printing and Content Delivery Lab at HP Labs, UK. He works in the Automatic Policy Enforcement eXchange Project with interests in information processing technologies; in particular, document workflows, data leak prevention, data mining and applications. He completed his PhD in Computer Vision at Cardiff University and HP Labs, and holds a MSc. in Mathematics and Computing from Swansea University, UK and a BSc in Mathematics from Cardiff University.

Dr. Helen Balinsky is a Principal Research Scientist from HP Labs, UK and is the Technical Lead for Automatic Policy Enforcement eXchange Project in Printing and Content Delivery Laboratory. Her research interests include: document security within an organization and in cross-organizational workflows, Data Leak Prevention, data-mining and security applications. Helen received her Master and PhD degrees in Applied and Industrial Mathematics from Technion –Israel Institute of Technology. Helen is a Fellow of The Institute of Mathematics and its Applications (IMA).