PDF security - a brief history of development

Background

Adobe was the first organization that set out to try and provide security controls for PDF based documents, and had their own particular views as to what users might (or might not) want in order to control the access to and use of information in PDF format.

Information security (in the sense of access controls, and continuing use controls) was not provided in earliest versions of PDF documents, simply because the most important feature of Adobe PDF was to ensure that what was shown on-screen or on a printed copy was the same, regardless of operating system, or printing device, being used.

PDF Password access controls

When Adobe first introduced PDF access control security, the controls the publisher selected were enforced by using passwords. Passwords were the commonest access control mechanism in use at the time, because, in fact, there was nothing else that was viable. But the way it was implemented was not a good idea, because it left it up to human beings to 'decide' what the passwords should be – and they inevitably chose passwords that were short and easy to cope with (and therefore easy for password crackers to attack) as against long, complex, and difficult to type in, because it was more important not to annoy your recipient than to worry about if what you were doing was realistically secure.

Unfortunately, using passwords as controls also allowed any recipient of a password protected document to pass it, and the associated password(s) to anyone they chose, and nobody was any the wiser. No mechanisms were created that could check that the person using the password was authorized to do that.

But even when the use of passwords was strengthened by implementing cryptography to prevent trivial access to the underlying document by simply decoding the PDF formatting, some fundamental weaknesses inherent in the use of passwords remained.

PDF Security and backwards compatibility

The first thing to be aware of is that the security applied to a PDF document is not simply a function of which version of the Adobe Writer/Viewer combination you happen to be using, but for backwards compatibility reasons, the features that were implemented in earlier versions have been carried forward to the very latest releases so as not to upset a large client base, and so your own requirements may not be exactly mirrored by the different Adobe products still in use.

For instance, if you go back to the security provisions of Adobe 5 (still highly popular, much implemented - especially by hackers because it had very weak controls as compared with Adobe 7 and later - but still able to process much of the files that Adobe 6,7, and 8 produced) you had two passwords, one (optional)

to be able to open the document, and the other (optional) to allow you to change the permissions (or limitations) that were applied to a document, and the presence of passwords.

The permissions that you could authorize, and therefore control, in Adobe 5 were:

- changing the document content;
- content copying or extraction;
- authoring comments and form fields;
- form fields fill-in or signing;
- content accessibility (using screen readers);
- document assembly;
- encryption level;
- printing (forbidden, low quality, high quality).

That was with the 'high' 128 bit encryption algorithm. Things were a little simpler if you used a weaker algorithm simply because you had fewer controls. But we are going to ignore this possibility.

The first thing to notice about the controls on offer is that they are unusual if you are trying to prevent uncontrolled circulation of a PDF formatted document. There is no concept of licensing, start and stop dates, control of numbers of views and prints, or identifying the licensed user.

Controlling the use of forms seems rather curious, if the purpose of a form is to have it filled in, and separating document assembly from content copying/extraction (where you could presumably do the same thing) does not seem immediately logical.

It is difficult, therefore, to reconcile the controls that were provided with what typical IPR owners normally want to control, when they provide their information to other people, especially when those people are not on their internal computer network and cannot be managed by controls other than those pertaining specifically to the document.

What controls do information sellers expect to have available?

Study has shown that the controls that publishers and corporate bodies actually want are much closer to the Digital Rights Management (DRM) controls that the music and film industries have been trying to implement, that would allow them to:

- prevent simple copying and redistribution of PDF documents;
- prevent simple Print Screen and screen capture programs;
- stop access to documents outside of a licensed period;
- limit the number of times a document can be viewed/printed (pay per view);
- if prints are made, be able to find who is distributing them by dynamically adding watermarks to screen images and printed pages;
- cease access to sensitive material as and when necessary, with confidence that the access will be ceased effectively.

These controls did not really align with the Adobe developed PDF controls, and, as a result, other PDF security providers began to emerge (see later). However, Adobe continued to control this space as the dominant supplier, even if they were not providing the DRM controls that publishers would have expected.

Later Adobe PDF security developments

By Adobe 7 things were different, in that Adobe had provided some slightly different options that provided the following controls:

- inserting, deleting and rotating pages;
- filling in form fields and signing existing signature fields;
- commenting, filling in form fields and signing existing signature fields;
- any except extracting pages;
- enabling copying of text, images and other content;
- printing (forbidden, low quality, high quality).

But these were also still being reconciled back into the permissions that Adobe 5 had provided because of the need for backwards compatibility.

The biggest change, apart from the update to the 128 bit AES algorithm (which made little practical difference because the biggest weakness of their implementation was actually an attack against the passwords protecting documents rather than an attack against the cryptography, which is far more difficult), was the addition of something called 'Encrypt for certain identities using digital certificates.'

It is necessary to discuss the concepts and realities of 'digital certificates' in order to understand what they are, and why they turned out to be useless as a PDF control mechanism.

Digital certificates and the great PKI fiasco

In the late 1990's and early 2000's the IT security industry developed a concept usually referred to as Public Key Infrastructure (PKI). The idea was that private companies would sell people 'digital identities' or 'digital certificates' that would uniquely identify them on the Internet and could make them personally liable for their activities (purchasing goods, paying bills, making government returns and so on).

The vast majority of normal people had never heard of such things as 'digital certificates', and most certainly did not have them. And given the (lack of) security about the Internet, there was actually every possible reason for people to make sure that they could not be claimed to be accountable, especially when they had no idea of how to protect themselves from things like identity theft.

It might seem strange to some, but consumers actually figured out that it is not a good plan to pay actual money to people in order to become personally liable for making payments, obtaining web site access, or anything else, when they can use credit cards which they can get for free and for which have little if any personal liability at all if anything goes wrong.

So as a result of this no-brainer the PKI industry imploded. People did not buy 'digital certificates' and so the IT industry could not use them as a means of applying controls.

Unfortunately Adobe had invested in the next range of PDF security by building upon the concept of 'digital certificates' and equally unfortunately that didn't happen, and so neither did the controls that were expected to be able to limit the use of controlled documents to defined people.

Lifecycle management

Another big concept being invested in at the same time was 'lifecycle management.' This concept came from the document management industries, who were interested in controlling the way in which documents were created, authorised, circulated and finally killed off.

Document management companies were very focused upon what happens inside of an enterprise, and how to control that, but had little or no focus upon what happened once documents went outside of enterprise control.

So whilst the Adobe type controls played well into internal document management, in the longer term they were less focused upon PDF controls in the broader market.

By now Adobe had published their document structures as standards through the Internet Engineering Task Force (IETF) and so any number of companies (including Microsoft) could readily convert their document formats to PDF at any stage. The practical effect of this was to separate the need for internal controls from the need to have external controls.

For Adobe this meant that although their publishing format provided a consistency across platforms and printers, it was less important to have internal controls for the PDF format because internal controls could be readily applied over other formats before converting to PDF for transmission to the final recipient. And, as has already been observed, Adobe based PDF controls were not aligned to market expectations.

Third party systems for protecting and revealing PDF information

Following the Adobe developments, a number of third party suppliers developed into this space. There were three broad groupings:

1 those continuing to support the Adobe controls but providing their own answer to replace the Adobe password access controls;

2 those providing DRM class controls, either inside the Adobe viewer or outside it;

3 those providing systems to break the Adobe controls by either revealing the control passwords or by providing plug-in components to the viewer that bypassed the Adobe security rules.

There are a large number of companies who have created a number of alternative approaches to hiding passwords from becoming visible (and therefore being able to make them longer and more resistant to password crackers). These can be supplemented by an Administration or permissions server which can register the use of a secret password and prevent its re-use.

However, all suppliers using these type of Adobe replacement access controls are likely to be exposed to the same weaknesses as Adobe itself (see later).

Suppliers providing their own DRM class controls are basically in two groups, those providing additional security whilst operating as a plug-in to the Adobe solution, and those who provide separately enforced controls.

Here there is a basic security trade between being able to guarantee to be 100% compatibility with any Adobe release (and the requirements for each release may differ and cause plug-in implementation problems) and being exposed to security weaknesses in Adobe implementations.

Those suppliers working as plug-ins are exposed to Adobe weaknesses and cannot remedy them, so all the attacks listed below may potentially be used against them.

Suppliers providing their own controls cannot be attacked through the Adobe weaknesses, and can be much more innovative about the granularity of DRM they can support than following the Adobe led model.

Additional controls such as dynamic revocation of access rights, the ability to control start and end dates for time limited material, the ability to allow identical access to groups of documents, are just a small number of examples of how PDF controls have been broadened in the developing market.

At the same time it is important to note that some companies have developed products that allow access to be regained to Adobe protected PDF documents.

Probably the most important of these was revealed through a ‘classical’ attack on the Adobe security mechanism carried out by the Russian company Elcomsoft. They attacked the mechanism that secured the passwords (granting uncontrolled access to PDF), allowing them to be revealed (their program was sold on the basis that it allowed document owners to recover passwords that they had forgotten, and that was obviously perfectly obvious and reasonable). The product was so good that the President of Elcomsoft was persuaded to visit the USA, and then arrested for prosecution under the Digital Millennium Copyright Act, which forbids anyone from investigating or circumventing security mechanisms! However, in the end he was not prosecuted.

The second was actually far more serious indeed. Adobe published statements to the effect that their controls, and those provided by their third party suppliers (using plug-ins), had been implemented on an ‘honour system’ – that they required that everyone who produced plug-ins to their product to obey the conventions of their security mechanisms, and not to defeat them.

Unfortunately, since piracy seems to be rampant, any reasonable analysis suggests that the ‘honour system’ cannot to be relied upon. And that is a fundamental flaw when it comes to security mechanisms. If you are going to

have security, then you simply can't rely on human beings to implement it. Passwords are weak because they can always be passed on, and all too often are chose by humans, and so can be 'guessed' by password crackers even if a mechanism as subtle as the password revealer of Elcomsoft is not available.

And where there is an honour system you will also find people who are without honour.

PDF security going forwards

In 2006 Adobe announced that they would be ceasing their own DRM type security offerings and also announced a new document format, based upon the Macromedia flash technology, following the bringing together of the two businesses.

However, the general publishing market (public and private) market is still very heavily reliant upon the PDF format for document publishing, and in many cases, not ready to use the flash technology because it is yet to be well established as a medium that can be relied upon to give the form and format that the publisher requires – not everyone wants their material presenting in the form of an eBook where it looks like a novel and the pages turn like a paperback. Magazines are presented in a different way from Witness statements to a Court, or to corporate press releases.

The PDF format is very well established, and it will be many years before it shows signs of waning, but the protection of PDF seems to have moved away from Adobe and into the newer companies providing their own controls.

PDF DRM as a standard?

Whilst there have been moves in some of the publishing industries to implement standards for DRM controls, these have received little interest amongst suppliers. Adobe has not bothered to look into implementing emerging industry standards, and no international standards have been promulgated so far.

Work has been carried out in the Document Management industry for standards over XML called MoReq2. However, since the vast majority of the publishing industry do not use XML to create and render their documents, it may be some considerable time before the worlds of content publication and content management move closer together.