Neutral Citation Number: [2018] EWHC 2442 (Pat)

Claim No: HP-2017-000019

IN THE HIGH COURT OF JUSTICE
BUSINESS AND PROPERTY COURT
OF ENGLAND AND WALES
INTELLECTUAL PROPERTY LIST (ChD)
PATENTS COURT

Royal Courts of Justice
Rolls Building
Fetter Lane, London, EC4A 1NL

Date: 28 September 2018

Before:

DAVID STONE
(sitting as a Deputy Judge of the Chancery Division)

- -

Between:

CLEARSWIFT LIMITED **Claimant**

- and –

GLASSWALL (IP) LIMITED

Defendant

- -
- -

Dr. Brian Nicholson and Mr. Christopher Hall
(instructed by CMS Cameron McKenna Nabarro Olswang LLP) for the Claimant
Mr. Richard Davis and Mr. Sam Carter
(instructed by Harbottle & Lewis LLP) for the Defendant

Trial dates: 10, 11, 12, 16 and 17 July 2018

- -

Judgment Approved

David Stone (sitting as a Deputy Judge of the Chancery Division):

1. These proceedings concern the validity of a single patent owned by the defendant, Glasswall (IP) Limited ("Glasswall"). The relevant patent is the United Kingdom designation of European Patent EP 1 891 571 B1 titled "resisting the spread of unwanted code and data" (the "Patent"). The Patent was applied for by an entity called Avecho claiming a priority date of 9 June 2005, and assigned to Glasswall prior to its grant in 2013.

2. The validity of the patent is challenged by the claimant, Clearswift Limited ("Clearswift"), an anti-malware provider, only on the basis of lack of inventive step over two items of prior art:

    (a) United States patent application 2005/0081057A1 entitled "Method and system for preventing exploiting an email message" ("Cohen") published on 14 April 2005; and

    (b) A series of posts on an Internet bulletin board under the subject line "Avecho Glasswall Anti virus technolog? [sic]" ("Avecho") published in full by 23 December 2003.

3. Clearswift seeks the revocation of the Patent under section 72(1)(a) of the Patents Act 1977 (the "Patents Act") for lack of an inventive step (sections 1(1)(b) and 3 of the Patents Act).

4. Although Glasswall initially asserted the independent validity of claims 2, 3 and 4 of the Patent, this assertion received very little attention during the trial and was dropped during closing speeches. There was no application to amend the Patent. There was also no counter-claim for infringement. Therefore, the only issue before the Court was the validity of claim 1 of the Patent over the two items of prior art.

5. Clearswift's counsel described claim 1 of the Patent as embodying "quite a simple concept". He also described this case in his closing submission as "relatively routine". Despite this, and the apparently narrow issue on invalidity, a significant amount of material was presented to the Court. Clearswift filed three reports from its expert (on the first day of the trial, I gave permission to adduce the last of these reports for reasons I set out at the time) totalling 628 paragraphs. Glasswall's two reports from its expert totalled 313 paragraphs. There were nearly 10 hours of cross-examination. In addition, I was presented with 27 technical papers totalling 528 pages (only 3 of which I was taken to during the trial) and a 711 page text book from 2005, *The Art of Computer Virus Research and Defense* by Peter Szor ("Szor"). In addition, there were a further 40 exhibits, only one of which I was taken to during the trial. Clearswift's skeleton arguments in opening and closing totalled 380 paragraphs, and Glasswall's 285 paragraphs. I have taken all of this material into account, but I have not referred to all of it in this judgment, because it is not necessary to do so. It was probably not necessary for much of it to be before the Court, particularly as this case was governed by the costs budgeting regime, limiting the recovery of costs.

6. From time to time, Clearswift's submissions appeared to suggest that the Patent was obvious over the common general knowledge. For example, in his opening submissions, Clearswift's counsel wrote "the Patent presents nothing more than a particular non-inventive, application of the well-established [common general knowledge] techniques". This was repeated in closing. This was not however, Clearswift's pleaded case, as expressly averred by Clearswift's counsel. I have therefore only assessed obviousness in relation to the two pleaded pieces of prior art.

7. I was made aware of unrelated proceedings between the parties in the United States involving different patents – I was asked to have no regard to those proceedings, and I have adopted that course.

8. Dr Brian Nicholson and Mr Christopher Hall appeared for Clearswift. Mr Richard Davis and Mr Sam Carter appeared for Glasswall.

Outline

9. Malware including computer viruses has been the scourge of computer users for many years. The rapid increase in Internet use (and particularly email use) in the 1990s led to an increase in the proliferation of malware, and a concomitant proliferation of anti-malware software to try to prevent it.

10. In 2005, the majority of anti-malware software worked by scanning for malware, or evidence of the presence of malware. Incoming files (including emails) were scanned, and blocked (or parts of them blocked) if evidence of malware was detected. Scanning systems could only look for what they knew – so new malware could not be detected until the anti-malware software had been updated. This was, in essence, a form of blacklisting – that is, known malware was included on the blacklist so that it could be stopped. The drawback was that malware could not be included on the blacklist until it had been detected, and the software updated.

11. Some anti-malware software also used whitelists – for example, allowing through files (or parts of files) including emails where the file type or the sender was recognised, and included on a safe list.

12. The Patent claims to take "an entirely different approach to protection against unwanted code". Rather than looking for known "bad" code, Glasswall says that the Patent teaches "parsing" and "regeneration" of each file – in effect, breaking it down, extracting its content and rebuilding it according to known rules – so that only "good" code is passed on to the recipient. Any "bad" code is left behind, but the regeneration occurs regardless of whether any malware is identified. The Patent also teaches a threat filter which allows through files which could not be parsed and/or regenerated if the sender and file type appeared on a whitelist.

13. Clearswift points to Cohen, a US patent published some 7 weeks prior to the Patent, and Avecho, a bulletin board exchange in which various experts discussed the Glasswall anti-malware software that was available at that time. Clearswift posits a different construction for the Patent, but says on any construction, there was no inventive step from Cohen and/or Avecho to the Patent. Clearswift says that Cohen and the Patent are doing the same thing in the same way. In relation to Avecho, Clearswift says that Avecho discloses to the skilled person the whole of the relevant claim of the Patent.

14. I am conscious that no brief outline can do justice to the detailed arguments advanced by the parties – the above is provided at the start of this judgment to provide context.

The witnesses

15. There were no witnesses of fact. Each party called a single expert witness.

Alexander Shipp

16. Alexander Lawrence Shipp is an experienced anti-malware practitioner. He is currently the Chief Technology Officer of Equine Register Limited, but in his earlier career, he developed expertise in designing and evaluating anti-malware technology. Mr Shipp studied computer science at Cambridge University and graduated in 1983. He worked in industry in a number of roles relating to computing before joining MessageLabs. MessageLabs offered a variety of anti-malware solutions. Mr Shipp was responsible for the team which developed MessageLabs' anti-malware capabilities. He had the title Imagineer – he was responsible for imagining things. He was a pioneer in the field of malware detection in real time. Mr Shipp also regularly visited and presented computer security briefings at organisations including GCHQ, NATO, the FBI and national Computer Emergency Response Teams in various countries. A list of Mr Shipp's ten patents was provided to the Court.

17. Mr Shipp was clearly a very knowledgeable witness. He was able to give well-considered evidence as to the common general knowledge of the skilled person, including as at the

relevant time. His evidence was criticised in two ways by Glasswall. These were said by Glasswall's counsel not to render Mr Shipp's evidence unreliable, but to "colour it". It was suggested that I needed occasionally to take care with Mr Shipp's evidence.

18. First, it was said that Mr Shipp could not reliably distinguish between what was known, and what was common general knowledge. It was also said that he could not reliably distinguish between what could be done and what was obvious to do. I accept this criticism to a limited degree. In my judgment, Mr Shipp well appreciated that he was not himself the skilled person for the purposes of the legislative test, and he did his best to ensure that his evidence was given from the correct perspective. I do not accept that he was "imagineering" in his evidence – a reference to his earlier job title put to him in cross-examination. However, as would be expected of an expert witness, he clearly knew more in 2005 (and knows more now) than the common general knowledge, and some care must be taken with his evidence in this regard.

19. The second criticism of Mr Shipp was that "he was there to argue his case". I accept this criticism to a very small degree.

20. These are, however, minor and understandable issues. Overall, I found him to be a witness who expressed himself clearly and was of significant assistance to the Court.

Professor Christopher Mitchell

21. Professor Christopher Mitchell is Professor of Computer Security at Royal Holloway, University of London, a position he has held since 1990. He has a BSc and PhD in mathematics. At Royal Holloway, he helped found the Information Security Group, a significant academic research body in information security. Additionally, Professor Mitchell has acted as a consultant on security matters, including as a member of Microsoft's Trustworthy Computing Academic Advisory Board from 2002 to 2014.

22. Professor Mitchell was a very knowledgeable witness who was able to explain the technology in an accessible fashion. He was open and frank in his replies to cross-examination.

23. Counsel for Clearswift did not criticise Professor Mitchell's expertise, but he did submit that Professor Mitchell was not the right person to assist the Court in this case. Unlike Mr Shipp, Professor Mitchell has never worked in industry, and has not in his academic career worked on or supervised anti-malware projects. Counsel for Glasswall conceded that Mr Shipp was therefore better placed to give first hand evidence of what was and was not common general knowledge in 2005, and what the skilled person would understand from the Patent, Cohen and Avecho. Counsel for Glasswall submitted, however, that Professor Mitchell had been able to educate himself from contemporaneous materials so as to be able to help the Court. Clearswift's counsel had no objection to this approach, but did submit that it was a difficult role to get right, and that those instructing Professor Mitchell had not got it right in this case. Counsel for Clearswift submitted that Professor Mitchell was unable to assist the Court from the correct perspective, which affected his (Professor Mitchell's) understanding of the Patent and limited his ability to envisage what someone working in the anti-malware industry would do without invention.

24. There is some force in these submissions, but they only go so far. In the end, the parties agreed on who the person skilled in the art is for the purposes of this case, and expressed themselves to be largely in agreement on the common general knowledge (although I return to this below). As Jacob LJ explained in *Technip France SA's Patent* [2004] RPC 46 at para 12:

> "I must explain why I think the attempt to approximate real people to the notional [person] is not helpful. It is to do with the function of expert witnesses in patent actions. Their primary function is to educate the court in the technology – they come as teachers, as makers of the mantle for the court to don. For that purpose it does not matter whether they do not approximate to the skilled [person]. What matters is how good they are at explaining things."

25. Both witnesses were good at "explaining things". Both were seeking to help the court. Both were able to fulfil the role of teacher. Neither is (or could be) an approximation for the skilled person – Mr Shipp because he is significantly too skilled and inventive, and Professor Mitchell because he has never worked in industry. But, as Jacob LJ set out, that is not their role. I am satisfied that both were able to assist the court and I am grateful to them for the time and effort they put in to help me understand the technology. It therefore also follows that I do not accept the various criticisms levelled during the trial at the way in which the experts were instructed.

26. It also means that where the experts disagree, it is not a matter of simply accepting one over the other because he was a better witness. The parties were agreed that all the relevant determinations (the construction of the Patent, the teachings of the prior art, and whether the Patent involves an inventive step) are, in the end, for the Court to make, based on the evidence before me. I have adopted that approach.

The law on the structured approach to obviousness

27. The parties agreed that the appropriate structured approach for me to take to the assessment of allegations of obviousness was that first set out by Oliver LJ in *Windsurfing International Inc v Tabur Marine (Great Britain) Limited* [1985] RPC 59 and restated by Jacob LJ in *Pozzoli SPA v BDMO SA* [2007] FSR 37 at [14] to [23] ("*Pozzoli*"):

> "(1)
>
> (a) Identify the notional "person skilled in the art";
>
> (b) Identify the relevant common general knowledge of that person;
>
> (2) Identify the inventive concept of the claim in question or if that cannot readily be done, construe it;
>
> (3) Identify what, if any, differences exist between the matter cited as forming part of 'the state of the art' and the inventive concept of the claim or the claim as construed;
>
> (4) Viewed without any knowledge of the alleged invention as claimed, do those differences constitute steps which would have been obvious to the person skilled in the art or do they require any degree of invention?"

28. I was also taken to comments by Floyd J (as he then was) in *Zipher Limited v Markem Systems Limited and Anor* [2008] EWHC 1379 (Pat) at para 284:

> "This approach assists the fact-finding tribunal, but is not a substitute for the statutory question: "is it obvious?" In applying it, as elsewhere, hindsight is impermissible. It has to be remembered that the skilled person is not in a position to perform his own *Pozzoli* analysis. It is particularly important to remember that the first three stages are merely those which the court needs to go through in order to equip itself with the tools to answer the statutory question, which is the fourth one. The first three steps involve knowledge of the invention, which must then be forgotten for the purposes of step 4. What one is seeking to establish is whether the claim extends to methods or objects which are, without knowledge of the invention and without inventive capacity, obvious."
>
> I have adopted that guidance.

29. It was common ground between the parties that I should undertake the various steps of the *Pozzoli* analysis as at the Priority Date. There was no challenge to the claimed priority date of the Patent of 9 June 2005. I was told that whilst the relevant date for determining what a piece of prior art teaches is the date of publication, it was conceded that nothing relevant happened between the publication of Avecho, the publication of Cohen and the Priority Date of the

Patent. Therefore, as agreed by the parties, whilst being mindful of the actual dates for undertaking these different tasks, it does not matter on the facts of this case.

The skilled person

30. In *Hospira UK Limited v Genentech* [2015] EWHC 1796 (Pat), Arnold J described the skilled person at para 29:

> "A patent specification is addressed to those likely to have a practical interest in the subject matter of the invention, and such persons are those with practical knowledge and experience of the kind of work in which the invention is intended to be used. The addressee comes to a reading of the specification with the common general knowledge of persons skilled in the relevant art, and he (or she) reads it knowing that its purpose is to describe and demarcate an invention. He is unimaginative and has no inventive capacity. In some cases, such as the present one, the patent may be addressed to a team of persons having different skills."

31. As Pumfrey J (as he then was) noted in *Conor Medsystems Inc v Angiotech Pharmaceuticals Inc* [2006] RPC 28 at para 35:

> "To an inappropriately defined skilled [person], nothing may be obvious or everything may be obvious. The most difficult part of any obviousness case is the attribution of the relevant skill and knowledge of the notional addressee of the patent. When the common general knowledge is identified, the height of the bar is set."

32. Although the skilled person was initially in dispute, by the end of the trial, the parties agreed that the skilled person is a person with a computer science degree and between one and three years' experience working in the anti-malware software industry.

33. There was some cross-examination as to whether the skilled person's degree would need to be a 2:1 or whether a 2:2 would be adequate. Counsel for Clearswift later conceded that it does not matter, and I agree. First, the skilled person is very diligent and is not forgetful: s/he will recall everything that has been taught in a degree course, even though a student will not. It is what is taught on the course that is relevant, not how any particular student will have performed in examinations. Further, as Counsel for Clearswift submitted, the skilled person is a notional legal construct. As Pumfrey J set out in *Halliburton Energy Services Inc v Smith International* [2005] EWHC 1623 at para 39:

> "The skilled person is essentially a legal construct, and not a mere lowest common denominator of all the persons engaged in the art at a particular time."

34. Counsel for Clearswift submitted that the legal construct of the skilled person is useful for several reasons: first, to ensure an objective assessment; second, to prevent particular experiences from tainting the common general knowledge and third, to remove the inventive potential that most real people have. I accept those submissions.

35. Counsel for Clearswift further submitted that, whilst there was eventually agreement as to the skilled person, because of initial lack of agreement, Professor Mitchell's evidence was given on the wrong basis. He had provided his evidence on the basis that the skilled person had a computer science degree, but limited, if any, industry experience. Counsel for Clearswift submitted that this did not matter for the purposes of the common general knowledge, because everything known to a computer science graduate will also be known to a computer science graduate with industry experience. However, in assessing obviousness, more will be obvious to the skilled person as ultimately agreed than will be obvious to the skilled person initially relied on by Glasswall. There is some force in this submission, and so I have taken care with Professor Mitchell's evidence on what would have been obvious to the skilled person under his definition. This does not mean that I must uncritically accept what Mr Shipp says about what would or would not have been obvious to the skilled person. As set out below,

obviousness is a multi-factorial assessment to be made by the tribunal on the basis of the evidence before it. Both parties agreed that, in the end, it is a matter for me.

36. I return below to the skilled person for the assessment of what the skilled person would have understood from the Patent, what they would have understood from the prior art, and how they would have developed the prior art without invention.

Common general knowledge

37. The common general knowledge is those matters which would generally be known and regarded as a good basis for further action by the bulk of those engaged in a particular art: *The General Tire & Rubber Company v The Firestone Tyre and Rubber Company Limited and Ors* [1972] RPC 457. The law on common general knowledge was not in dispute. I was taken to *KCI Licensing Inc v Smith & Nephew plc* [2010] EWHC 1487 (Pat) where Arnold J set out the law at paragraphs 104 to 112. That statement was approved by the Court of Appeal: [2010] EWCA Civ 1260. It is not necessary to excerpt those judgments here.

38. Whilst the common general knowledge was in dispute at the start of the trial, in their closing skeleton arguments, the parties summarised their positions on the common general knowledge as follows:

> Clearswift: "in relation to the written evidence, anything said to be [common general knowledge] by either expert, or contained in Szor, is part of the [common general knowledge]".
>
> Glasswall: "we do not understand the parties to be in any substantial disagreement on this point, certainly not to the extent that anything needs to be resolved."

39. During closing speeches, I expressed some difficulty with the parties' approach. If the parties were agreed as to the common general knowledge, I invited them to set out the terms of that agreement. There would be little point to my summarising thousands of pages of evidence (including over 700 pages of Szor) if my summary was not what was ultimately agreed between the parties. Each side then submitted, after the trial had closed, a summary document which they averred referred me to the relevant parts of the evidence, helpfully collected under headings. The two documents overlapped in large degree, albeit with differences of emphasis.

40. In the end, the issues between the parties were comparatively narrow. It does seem to me that more progress could have been made prior to trial on reaching agreement as to the skilled person and the common general knowledge, with savings of court time and costs for both parties. With hindsight, it is easy to suggest that a mandated meeting of the experts and/or a pre-trial review (neither of which occurred in this case) may have assisted even if only to focus the parties' attentions earlier.

41. I turn now to my findings of the common general knowledge as at June 2005. The Patent is concerned with preventing the spread of unwanted code and data in files and emails. I have therefore been asked by the parties to set out the common general knowledge in relation to files, emails and attachments, as well as malware and anti-malware systems, before turning to the key issues of parsing and regeneration and threat filters. In the end, little turned on much of this, so I have summarised or abbreviated the parties' submissions where possible. Although the below is expressed in the present tense, it sets out the common general knowledge as at June 2005. The headings are those provided by the parties.

Files and formats

42. Computer data are typically organised in discrete files, each with a specific purpose, stored in a readable format. Files may contain, for example, software, documents or images. All files generally have some sort of file format: file formats are used to specify exactly how stored information should be recorded and interpreted within the file. Some file formats are comparatively simple – others involve very complex data structures. Common computer file formats include Microsoft Word documents and PDF files. The file format defines the way in

which information is arranged in an object or container, and is necessary to enable an operating system to read a file to extract the information. As is apparent from the examples above, file names are generally used to indicate file types.

43. Both experts agreed that "file" is an ambiguous term. Its meaning can therefore depend on context. Arranging data into a particular format so as to create a file is referred to as "writing", "composing", "generating" or "creating".

Specific file types

44. Some files have complicated, proprietary file formats, such as Microsoft Word 2003 (.doc) – other files have open source formats (such as JPEG files and CSV files). Some file formats are the subject of international standards, such as PDF files and HTML files. There is no single method by which the format and nature of a file can be detected automatically. There are a huge number of file formats, some very obscure, and more are being created all the time.

45. Encryption of a file obfuscates the file data, and may prevent recognition of the file type.

Emails

46. The Internet-based system used for sending, relaying and receiving emails is built upon a few simple and well-established building blocks. The Simple Mail Transfer Protocol (SMTP) is a set of rules for defining how two machines communicate. SMTP only allows the exchange of text-based messages, where these text-based messages can only contain characters used in normal printed text, such as letters, digits and punctuation.

47. An email is, in essence, a collection of 0s and 1s, combined according to a set of rules so that the message can be interpreted. One well known set of rules for combining 0s and 1s into basic characters is known as the American Standard Code for Information Interchange (ASCII). The ASCII character set represents characters in 8-bit binary format. The original ASCII character set was limited to 128 characters, including the Latin alphabet, Arabic numerals, some punctuation marks, and certain control characters, such as backspace and horizontal tab. An extended form of the ASCII character set was introduced in the 1980s, and included a further 128 characters.

48. The rules for combining ASCII characters into an email so that it can be read and interpreted by the receiving computer are set out in a protocol known as the Internet Message Format (IMF). The IMF is set out in standards known (for historical reasons) as "Requests for Comment" (RFCs). The current RFC 2822 dates back to 2001, and states:

> "…a message is a series of characters. … Messages are divided into lines of characters. A line is a series of characters that is delimited with the two characters carriage-return and line-feed; that is, the carriage return (CR) character (ASCII value 13) followed immediately by the line feed (LF) character (ASCII value 10). … A message consists of header fields (collectively called 'the header of the message') followed, optionally, by a body. The header is a sequence of lines of characters with special syntax as defined in this standard. The body is simply a sequence of characters that follows the header and is separated from the header by an empty line (ie, a line with nothing preceding the CRLF).
>
> …
>
> Header fields are lines composed of a field name, followed by a colon (":"), followed by a field body, and terminated by the CRLF. A field name MUST be composed of printable US-ASCII characters (ie, characters that have values between 33 and 126, inclusive), except colon. A field body may be composed of any US-ASCII characters, except for CR and LF."

49. Thus, a message transferred using SMTP consists of a header (containing a sequence of header fields) and a body (containing the message). Common header fields include to, from,

cc, bcc etc. An SMTP email has a number of disadvantages – it can use only the designated ASCII characters (excluding non-Latin alphabets or writing systems); bold or underlined text is not possible; and nor can files be attached.

50. These perceived failings in SMTP were remedied in approximately 1992 with the introduction of Multipurpose Internet Mail Extensions (MIME). MIME, defined by a number of RFCs which have been revised over the years, allows richer email content. MIME-encoded messages have a similar overall format to that defined in RFC 2822 – a number of header fields and a single body.

51. In a MIME email, a MIME header is added to the overall email header, and each individual MIME part has its own header, followed by a blank line, followed by the content of that part, and terminated by the MIME boundary. Encoding in this way allows attachments to be represented in MIME format as separate sections of the email. Those sections could then be decoded by the receiving email user so that the attachments could be opened and read.

52. A simple example of a MIME message is available on the Microsoft website:

```
From: John Doe <example@example.com>
MIME-Version: 1.0
Content-Type: multipart/mixed;
        boundary="XXXXboundary text"

This is a multipart message in MIME format.

--XXXXboundary text
Content-Type: text/plain

this is the body text

--XXXXboundary text
Content-Type: text/plain;
Content-Disposition: attachment;
        filename="test.txt"

this is the attachment text

--XXXXboundary text--
```

Malware and exploits

53. Malware is any software which, when present on a computer, has the capacity to perform actions on that device without the consent of the owner. Malware is a portmanteau word constructed from "malicious software".

54. Computer viruses are software programs that self-replicate. They are regarded as a particular category of the more general notion of malware. A virus is embedded in another program or file – when the host program or file is run, the virus program is also run. The virus spreads by self-replicating, copying its code to be embedded in further programs or files. Viruses have also been developed to become polymorphic, enabling them to self-alter their form to avoid detection.

55. There are many types of computer viruses. One type is called a macro virus. Macros are code, written in a scripting language, that automate useful tasks within documents. The scripting language can also be used to create self-replicating code including viruses. The popularity of Microsoft Office documents, which use macros, and the comparatively low skill level needed to write a macro virus, meant that macro viruses became very common during the 1990s.

56. It is also important to explain the notion of exploits. Exploits are not malware as such, but rather software vulnerabilities which provide a method for delivering malware or allowing it to run. Exploits include:

    (a) Buffer overruns: When a program runs it writes data to a size-limited buffer. A buffer overrun is an anomaly whereby the program overruns or overflows the buffer and gains access to adjacent memory space. A buffer overrun can allow malware to run. One way to create a buffer overrun is to send a malformed email, or an email with malformed headers. Certain values within Word files can also cause buffer overruns. The buffer overrun can be exploited to enable the virus – for example, a virus can be included in an image file by including the "bad" data after the "official" end of the file, and can be triggered by a buffer overrun elsewhere in the file.

    (b) Malformed MIME: The MIME protocols are open to a degree of interpretation. That difference in interpretation can be exploited and used to hide a virus. For example, a malformed header could be exploited.

    (c) Encoding format nuances: Nuances in various encoding formats could also be exploited. This difference in understanding of encoding formats can be exploited to spread a virus, for example, encoded as an attachment.

    (d) File-type masquerading: This is done, in effect, by the malware "masquerading" as a different type of file. The malware follows the naming convention of an email attachment considered to be safe (for example, picture.bmp) and describes the file in the relevant headers as being a BMP file. In reality, the file is an unsafe executable, which runs when the user attempts to open what is perceived as being a safe file.

Knowledge of anti-malware techniques

57. From the late 1980s onwards, rules, procedures and computer programs were devised to try to address the threat arising from malware. This gave rise to the general notion of anti-virus (AV) or anti-malware software. Anti-malware software can potentially function in a range of ways, although most commonly it is used to scan a computer by looking through some or all of the files stored on that computer and trying to detect the presence of malware in these files. This is conventionally achieved by looking for known sequences of bytes in a file that characterise a particular (known) type of malware, so called virus signatures.

58. The major drawback of such signature scanning is that it can only detect known viruses. This requires anti-malware software providers to identify new viruses very rapidly indeed, and then provide updates to users.

59. Owing to the rapid increase in malware, the anti-malware market in the early 2000s was moving very rapidly. By that time, pure signature based scanning was considered insufficient, and all major anti-malware companies included some form of heuristics (discussed further below) in their products. In essence, whilst most anti-malware products used a battery of techniques, they operated largely through detecting viruses by looking for strings of dangerous code, and evidence of virus behaviour.

60. These issues, and the best approach, were hotly debated in the anti-malware industry in the early 2000s. One school of thought put it (colourfully) like this: signature detection is akin to people leaving open the front door to their home but denying entry only to known burglars – a better approach would be to close all the doors and only open them to known guests. These views were widely disseminated.

61. In addition to the change in detection techniques, the need to prevent email transmission of malware led to a convergence of technology between the fields of anti-malware, anti-spam and email and policy filtering. A policy engine would typically be positioned at a corporate gateway to enforce various policy rules, including, for example, on attachment size. By the

early 2000s, there was regular interplay between malware detection products and policy engines.

Specific anti-malware techniques

62. Other anti-malware techniques in use include:

(a) Firewalls: Firewalls operate by trying to prevent malware from ever reaching computers, including by scanning traffic on the computer network. There are many types of firewall, but they all scan traffic for content which is disallowed and then prevent that content from passing through.

(b) Whitelisting: Whitelisting involves the blocking of any content which has not been pre-identified as being safe. Any file type which is known to be safe is included on the whitelist. When that file type is identified by the anti-malware engine, the file is presumed to be safe. Whilst whitelisting is effective in blocking malware (so long as the whitelist is accurate), the disadvantage is that it blocks large amounts of content which is safe, but has not (yet) been included on the whitelist. These are known as "false positives".

(c) Blacklisting: Blacklisting is the opposite of whitelisting – it blocks only content which is known to be unsafe. Blacklisting is also done by file type. Blacklisting carries the risk that unsafe content which has not been listed will be treated as safe (known as "false negatives"). More anti-malware software uses blacklisting than uses whitelisting.

(d) Heuristics: Heuristic detection analyses files or parts of files for unexpected or questionable format or content. Heuristic techniques include code emulation and the use of machine learning techniques to try to identify viruses by looking at their behaviour. Commonly known techniques included:

i. Disassembly and evaluation of source code;

ii. Emulation (similar to sandboxing, discussed below);

iii. Positive feature detection (identifying content or features known to be malicious);

iv. Negative feature detection (identifying content or features known to be safe);

v. Geometric detection (identifying variation from file specifications); and

vi. Analysis of known areas of exploitation, such as header content and size and redirections.

(e) Sandboxing: Sandboxing is a process whereby a computer program is allowed to run in a virtual or highly restricted environment from which it cannot access critical systems.

(f) Macro removal: As set out above, macros are well known as particularly common vehicles for viruses. Macro-removal is a process whereby anti-malware programs are able to remove macros from documents.

Content filtering

63. One further widely used scanning technique not discussed above is what is referred to as a filter or content filter. A content filter is a software filter that allows administrators to restrict accessible content from within a network. Email filters attempt to check the content of email messages before they are opened by an email program. Filters involve scanning for disallowed content. Filters are highly configurable, and can filter for attachment type or size, block inappropriate images or profanities, and strip emails or attachments of unwanted content. Content filters can scan for keywords in emails and in any attachments.

64. Having identified unwanted content, filters deal with it in two main ways – the unwanted content can be blocked, or cleaned to remove the malware. Blocking is comparatively simple – the anti-malware engine simply deletes the file so that it (and therefore the malware) is never delivered. Cleaning is more complicated. This requires the data to be rewritten in such a way that they no longer include the unwanted content. This is done in various ways:

    (a) Some anti-malware products blanket ban certain forms of files – these are simply removed from emails.

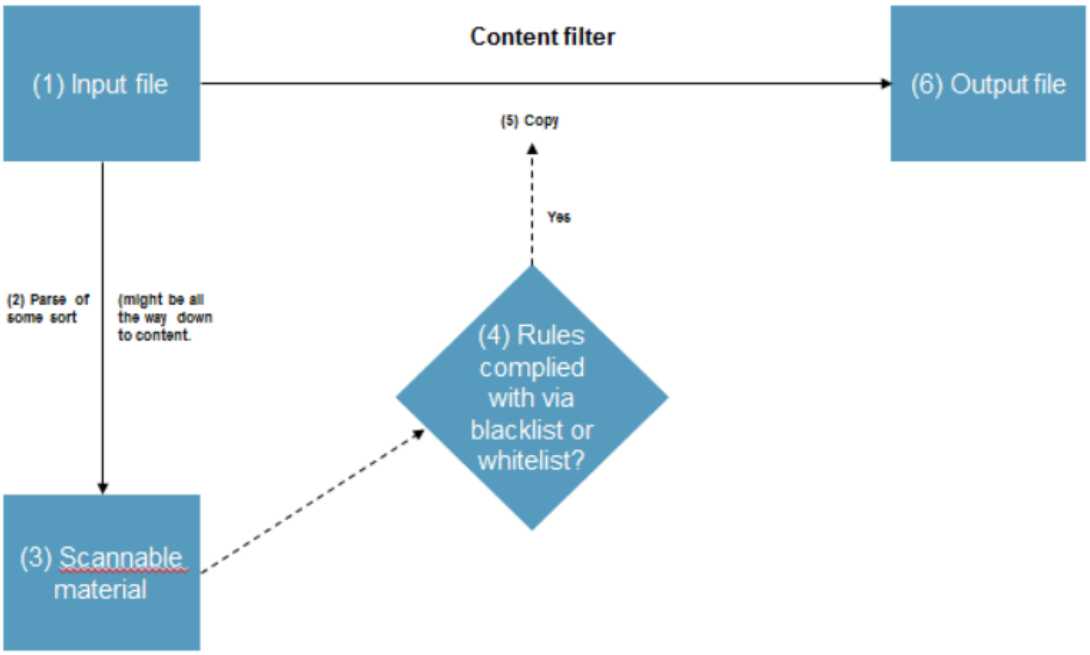
    (b) More sophisticated anti-malware products only ban those parts of files that are considered dangerous – for example, stripping macros from Microsoft Office documents.

    (c) Where a string of code is believed to be malware, some products remove the entire string of code, whereas others remove only the "pointers" at the start of the code, so that the malware is still present but cannot be executed.

65. If some unwanted content is cleansed from an email, the unwanted part is often replaced by the anti-malware engine with a notification, so that the recipient knows that something is missing. An example would be "The attachment has been removed". Anti-malware engines also routinely add different banners to the beginning or the end of incoming and outgoing emails that are deemed virus-free, to notify the recipient that they are considered safe. An example would be "This email has been scanned by xyz product".

66. MIME emails can be checked and cleaned/rewritten. To analyse the email, the anti-malware program "takes the email to bits" by parsing the email and identifying the parts with reference to the RFCs. Parsing enables a scan to be performed. A basic content filter goes part by part and makes an accept/reject decision in relation to each part of the email. In general, if a part is accepted, it will be passed to the output in the form received. There are exceptions to this, including, for example, the addition of the "This email has been scanned" wording set out above. If an attachment is removed, the MIME header for that attachment will be rewritten for the new text message "The attachment has been removed". The passed parts of the email are then recomposed in accordance with the RFCs and MIME protocols.

67. The common general knowledge on content filters was summarised by Glasswall's counsel in the following diagram.



68. As can be seen, the filter operates by parsing a file, and then either passing on the parsed content, or not.

69. Macros can be stripped in various ways. Some programs remove all macros – this is unacceptable for some parts of businesses that rely on macros (including, for example, in accounting spreadsheets). Other programs whitelist certain macros and remove others. Other programs remove macros if malware is detected, either by removing the whole macro, or by removing the part of the macro believed to be malware. Stripping out macros requires an understanding of the file format, or at least of the part of it being modified.

Bypassing anti-malware systems

70. As noted above, heuristic methods increase the number of false positives, preventing legitimate content from reaching its destination. One solution to this is to whitelist certain types of files or parts of files (for example, removing macros also reduces functionality of documents, so certain macros can be whitelisted, whilst all others are removed.)

71. Another solution is to whitelist by sender – thus, if the data come from a trusted sender, they are allowed through the anti-malware program without checking.

72. By June 2005, it was known to whitelist by file type OR by sender – but whitelisting by file type AND sender was not known.

Parsing and the converse

73. Parsing is not a term of art. Mr Shipp's definition of parsing was "taking the file and reading it in a logical order". Professor Mitchell's definitions of parsing were "disassembling an input sequence of symbols, assumed to be in a format defined according to known rules, into its constituent parts" and "analysing (a string or text) into logical syntactic components". I do not consider that there is any relevant difference between the definitions. If there is a difference, it does not matter for present purposes.

74. It was agreed between the parties that parsing requires rules. Parsing is only capable of processing information that is structured in accordance with the specification. How parsing is carried out will depend on the purpose of the parse, or, to put that another way, the extent to which one parses a file depends on what one is looking for and what one is trying to do. Thus, depending on the purpose of the parse, it may not be necessary to understand the file format in full.

75. There was some disagreement between the parties on the difference between scanning and parsing. I accept that these are not completely independent processes – and that to scan data, one would parse it first.

76. Parsers can be very simple – undergraduate computer science students will write parsers as part of their course. But parsers for complex files can be very difficult to write. Parsers are also available "off the shelf".

77. Creating, or constructing, a file is the reverse of parsing – the program uses its syntactic understanding of the relevant information to construct a file in the appropriate format which can be correctly parsed by another program.

78. This was the common general knowledge in June 2005.

The Patent

79. Following the *Pozzoli* stages, having identified the skilled person and the common general knowledge, I must now identify the inventive concept of the claim in question or, if that cannot readily be done, construe it.

80. The Patent first discusses the background art (Clearswift says inadequately) and explains the concept of computer malware. It summarises the two main approaches to virus detection:

    (a) Scanning incoming files for a virus signature; and

    (b) Scanning for evidence of virus behaviour.

81. At [0010] the Patent notes that it "takes an entirely different approach to protection against unwanted code".

82. The parsing/regeneration process for which Glasswall contends is then described at [0013]. At [0015] the Patent notes "the substitute file [ie, that which would ultimately reach the user] will be generated using a generator routine which can generate only 'clean' code data. It is therefore incapable of generating unwanted code matching any code in a received file." At [0022] the Patent repeats that, rather than detecting viruses or virus like behaviour, the engine "substitutes" files being "generated" which cannot contain unwanted code. At [0026] the Patent indicates that what passes through the system is not the original file, but a substitution. The threat filter is mentioned at [0028] and [0029].

83. The description of a first embodiment is at [0032] to [0051]. In this embodiment, there is both file format checking [0038] and file content checking [0039]. It is the parser that checks for file structure conformity: [0040] and [0041]. The process of parsing and content extraction is explained at [0043]. If the file format is recognised, it will be parsed. Data are then extracted from it, and temporarily stored in a data structure. A substitute file can be regenerated from the content data. [0048] deals with the parsing of parts of files.

84. The threat filter is described at [0050] and [0051]. If, on parsing, a file is not recognised, it is passed to the threat filter, which may let it through on the basis of its file type AND its source. Otherwise it is blocked.

85. The second embodiment, which relates specifically to emails, is at [0052] to [0100]. The engine parses the incoming email [0056]: each part of the email that conforms with a pre-determined data format is regenerated. The data type may be determined through analysis of both the file header and the body of the file [0059]. The parse/regenerate process may

therefore use a large number of separate conformity analysing devices, such as ASCII, CSV, RTF, TIFF, depending on the content of the email. Each of these parses the data according to the rules, and checks that the data correspond to the data format. Conforming data are "extracted" [0057], [0089] and "regenerated". Non-conforming data are blocked, or regenerated in a way which does conform.

86. [0092] establishes that only those parts of the file which conform with the pre-determined data format are regenerated and passed to the next stage. Any parts containing malware would not conform, and would therefore not pass (they would be blocked).

87. The third embodiment is set out at [0101] to [0111], incorporating the features of the second embodiment, but giving more detail in relation to the threat filter. If the part is not a permissible file type AND from a permissible source, it is blocked [0106]. If it is "whitelisted" by file type AND source, the part is let through and reassembled in the substitute email [0107].

The claim

88. It is common ground that the only claim which it is necessary for me to consider is claim 1 of the Patent. I have set out claim 1 below, broken down into integers (as the parties did), inserting headings (which were not contested) and with obvious typographical errors corrected, but I have reminded myself that the headings and integers are merely a convenient way to work through the claim. It is the claim I must construe.

"BACKGROUND

[1A] A computer implemented method of resisting the spread of unwanted code and data in an electronic file, the method comprising:

[1B] receiving an incoming electronic file wherein the incoming electronic file is an email having plural parts from a sender,

[1C] each part of said file containing content data in a pre-determined data file type,

THE RULES

[1d] each data file type having an associated set of rules;

[1e] said rules including the rules making up the file type specification

[1f] and additional rules constraining the values and/or ranges that content and parameters can take on

ANALYSIS

[1f1] determining a purported predetermined data file type of each part

[1g] parsing the content data of each part in accordance with the rules associated with the purported predetermined data file type;

[1h] determining if the content data of each part does conform to the rules associated with purported predetermined data file type;

REGENERATION & BLOCKING

[1i] regenerating the conforming parts of parsed content data,

[1j] upon a positive determination from the determination means, to create a substitute regenerated electronic file in the purported pre-determined data file type,

[1k] said substitute regenerated electronic file containing the regenerated content data;

[1l] blocking the parts of the parsed content data that do not conform to the rules associated with the purported predetermined data file type so as to block them from inclusion in the substitute regenerated electronic file,

THREAT FILTER

[1m] storing a list of file types and sources associated with said file types that are not considered a threat;

[1n] forwarding the non-conforming parts to a threat filter;

[1o] determining by the threat filter for each non-conforming part whether that non-conforming part is to be allowed through on the basis of the stored list and the sender of the file and the data file type; and

[1p] allowing a non-conforming part to bypass the blocking and including the bypassing non-conforming part in the substitute regenerated electronic file it determined to be allowable."

Inventive concept

89. There was no agreement between the parties on the inventive concept. Counsel for Clearswift submitted that the fundamental concept of the Patent is "actually very straightforward indeed" – it is to reduce the risk that files passing through it can be used to exercise exploits. It does this by allowing the skilled person to define a bespoke predetermined format to specify what is a "'normal', acceptable file" and then allowing only such files as comply with that

predetermined format to be passed through. Clearswift submitted that the Patent sets out at least two different techniques by which the rules can be used to ensure conformity:

(a) A "correcting" system based on a "loose reader" and "tight writer". Clearswift described the fundamental characteristics of this system as being that (i) it can correct non-conformities within parts of a file and (ii) it uses different rule sets for reading and writing; and/or

(b) A "non-correcting" system that uses only a single rule set to form a "conformant parser" that can only check that each part fully conforms with the defined rule set. This, said Clearswift's counsel, can be used with a regenerator that only amends parts to do the necessary updating (of headers and the like) when non-conforming parts are omitted (that is, blocked).

90. Glasswall contended that the inventive concept is that each part of incoming emails/files is parsed so as to extract its contents. Substitute parts are then regenerated from those contents. Importantly, this process of the regeneration of parts occurs regardless of whether any malware or other malformation is detected. This process of parsing, content-extraction and regeneration of that content *of itself* is said to provide a useful improvement in the elimination of malware.

91. In addition, the Patent includes a self-contained threat filter. The parties agreed that the threat filter is a sender AND file type whitelist. Glasswall submitted that the Patent provides for that whitelist to operate only on files that could not be parsed and regenerated, and that that had value to users. Clearswift rejected this contention.

92. In *Pozzoli*, Jacob LJ said at para 19:

> "In some cases the parties cannot agree on what the concept is. If one is not careful such a disagreement can develop into an unnecessary satellite debate. In the end what matters is/are the difference(s) between what is claimed and the prior art. It is those differences which form the 'step' to be considered at stage (4). So if a disagreement about the inventive concept of a claim starts getting too involved, the sensible way to proceed is to forget it and simply to work on the features of the claim."

93. I have adopted that approach, and turn now to construing the Patent.

The law on patent construction

94. The law on claim construction was not in issue between the parties. I was referred to section 125 of the Patents Act and to *Saab Seaeye Limited v Atlas Elektronik GmbH and Anor* [2017] EWCA Civ 2175 and particularly to paragraphs 18 and 19 of that decision. It was conceded that no issue of equivalents arises in this case. I have therefore taken into account the matters set out in paragraph 18 of *Saab v Atlas* citing the Court of Appeal in *Virgin Atlantic Airways Limited v Premium Aircraft Interiors UK Limited* [2009] EWCA Civ 1062 at para 5. I have not excerpted those well-known passages here.

95. I was also taken to Pumfrey J's judgment in *Halliburton Energy Services Inc v Smith International (North Sea) Limited* [2005] EWHC 1623 (Pat) where he said (at para 69):

> "I would diffidently add three observations of my own. The first is merely the trite principle that the addressee of the specification is the person skilled in the art, who approaches the document with the common general knowledge. Second, there may be obscurities and difficulties in a claim that cannot be resolved by an appeal to context. It is very rare that some sensible meaning cannot be attributed to the words used in a patent claim, but where a claim permits alternative interpretations it is possible to be left with no alternative but to take the most straightforward. Finally, and most importantly, over-meticulousness is not to be equated to

> carefulness. Care in working out what the patentee was aiming at when he chose the words he used is absolutely necessary."

96. My attention was drawn to the following propositions:

(a) A patentee is not entitled to use the specification either to write-out a limitation in the claim, or to write-in a limitation to the claim;

(b) Specification and claims have different functions. As Laddie J said in *Merck & Co Inc v Generics (UK) Limited* [2004] RPC 31 at para 38:

> "The purpose of the patent is to convey to the public what the patentee considers to be his invention and what monopoly he has chosen to obtain. These are not necessarily the same. The former is primarily to be found in the specification and the latter is primarily to be found in the claims".

Clearswift submitted that this approach applies equally in respect of exemplary implementations used in the specification to illustrate a patentee's purported invention;

(c) There is no presumption that a patentee has claimed the full breadth of all s/he teaches; and

(d) Although the claims are to be read purposively, it is not legitimate either to cut down or to extend the clear meaning of the language of a claim by reference to the body, for that would be to amend the claims. As Floyd J (as he then was) said in *Nokia v IPCom* [2009] EWHC 3482 (Pat) at para 41:

> "Where a patentee has used general language in a claim, but has described the invention by reference to a specific embodiment, it is not normally legitimate to write limitation into the claim corresponding to details of the specific embodiment, if the patentee has chosen not to do so. The specific embodiments are merely examples of what is claimed as the invention, and are often expressly, although superfluously, stated not to be 'limiting'. There is no general principle which requires the court to assume that the patentee intended to claim the most sophisticated embodiment of the invention. The skilled person understands that, in the claim, the patentee is stating the limits of the monopoly which it claims, not seeking to describe every detail of the manifold ways in which the invention may be put into effect."

I accept these propositions and have applied them.

97. I also note the position, agreed by the parties, that "parsing", "regeneration", "rewriting" and "stripping" are not terms of art. These terms therefore need to be understood in context.

Clearswift's construction

98. I was told by both parties that construction of the Patent was ultimately a matter for me. But I was also cautioned in strong terms of the need to view the Patent through the eyes of the skilled person. I have done so, relying on the expert evidence before the Court.

99. Clearswift made a number of criticisms of the drafting of the Patent, including its failure properly to acknowledge the prior art. I do not consider there to be anything helpful in these criticisms. Nor do I consider that it matters whether or not the system set out in the Patent was "100% effective".

100. As noted at the opening of this judgment, Clearswift's counsel described claim 1 of the Patent as embodying "quite a simple concept". Clearswift submitted that the Patent is about "appreciating that malware can exploit vulnerabilities and potential vulnerabilities in down-

stream software whose parser has been written to expect only properly formatted files". The Patent reduces the risk that files passing through it can be used to exercise exploits.

101. Clearswift's summary of its construction of claim 1 was as follows:

> "a. The claim relates to a method for an [antivirus] application that operates on an incoming electronic file that is an email and which has plural parts (integers a and b), each such part having a pre-determined data file type (integer c).
>
> b. Integers d to f require that each data file type has an associated set of rules, which is made up of two parts:
>
> i. rules making up the file type specification (integer e); and
>
> ii. additional rules limiting the values and/or ranges that content and parameters can take (integer f);
>
> c. reproducing each part only if it conforms with the associated set of rules (integers f1 to l); and
>
> d. forwarding non-conforming parts to the threat filter (integers m-p) that can nevertheless allow reproduction of the part if the email has been sent from a pre-approved sender (referred to as a source)".

102. Whilst Clearswift's counsel made detailed submissions on the interpretation of "parts" and "sub-parts" as used in claim 1, he conceded that ultimately it does not matter. I agree. "Part" and "sub-part" are not terms of art. Given a purposive construction in context, "part" simply means less than the whole. Clearswift's counsel conceded that all the claim requires is that each part has a data file type.

103. Clearswift's counsel also submitted that the set of rules associated with each data file type is made up of two sub-sets of rules – the first being the file-type specification (or at least some of the rules in the file type specification) and the second being additional rules being those used by the skilled person so as to delimit "'normal', acceptable file[s]". Examples of these were given by the experts – for example, ensuring the date field of an email contains fewer than 100 characters.

Glasswall's construction

104. Although Clearswift divided claim 1 of the Patent into integers, Glasswall submitted that it was more helpful to divide it instead into five "features". The headings for these five features are set out above. For my part, I found the feature analysis of more assistance than the integer analysis, although, as I have stated earlier, I have construed the claim on the basis of the totality of the words as set out in the claim.

105. Glasswall construed the claim as follows:

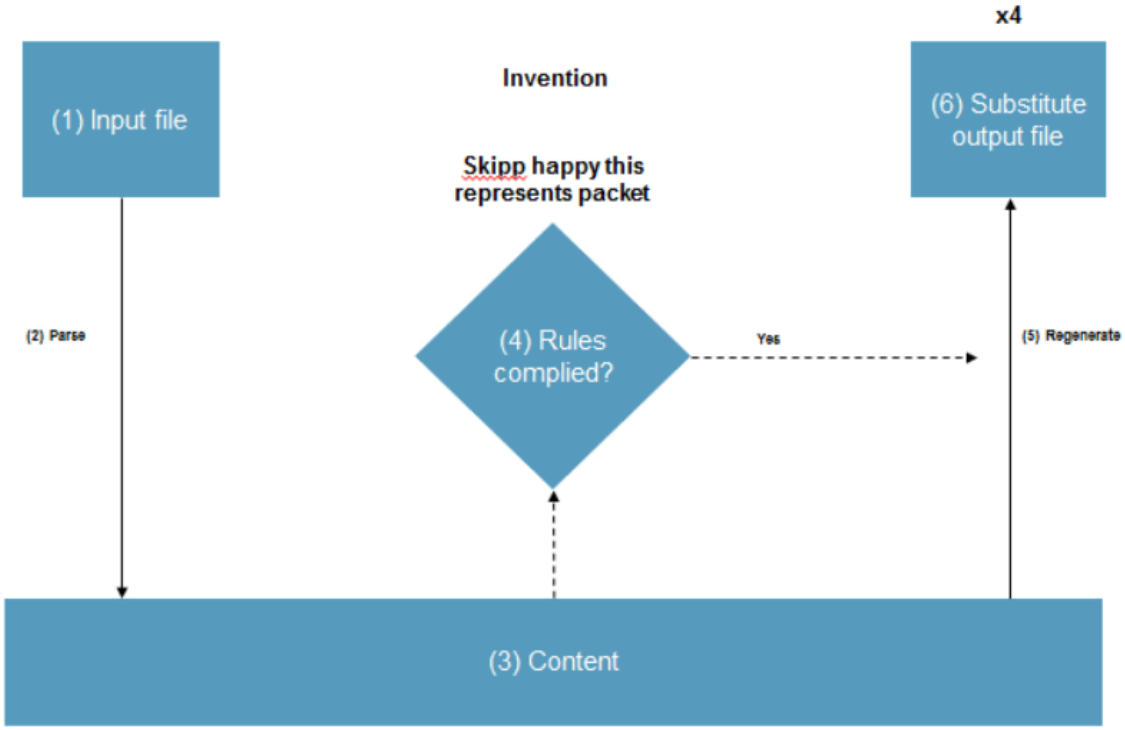
> (a) The Background feature establishes that claim 1 is a method claim, is computer implemented and operates on incoming emails, each part of which contains data in a pre-determined data file type. Importantly, it defines the email as having plural parts, and that each part has its own data file type.
>
> (b) Next, the Rules feature establishes what is meant by a "pre-determined data file type", and the two sets of rules associated with that data files type: first, the rules which make up the file type specification and second, additional rules constraining content or parameters.
>
> (c) The Analysis feature describes the first process set out in the method: analysis. For each part of the email, first, the predetermined data file type is determined, then the part is parsed using the rules associated with the predetermined data file type, and third, each parsed part is then checked for conformity with the predetermined data file

type. Here, parsing means that the content is extracted and temporarily stored in a data structure.

(d) The Regeneration and Blocking feature describes the second process set out in the method: the processing of each parsed part of the incoming email according to the results of the earlier analysis to create a substitute file containing the processed data. This is done by regenerating conforming parsed parts from the temporary data structure holding the parsed parts into a new substitute file. Regeneration requires construction of a new substitute file from the various conforming and regenerated parsed parts according to the rules which make up the relevant pre-determined data file type. Any non-conforming parsed parts are blocked, that is, they are not regenerated and are not therefore included in the substitute regenerated file.

(e) The Threat Filter describes the third process performed under the method – a mechanism by which non-conforming parsed parts may be included in the substitute regenerated file if they are determined not to be a threat. This is done under the method by forwarding the non-conforming part to a threat filter. The threat filter analyses the part on the basis of a stored list of permissible data file types AND sources. If permissible, the non-conforming part is included (unregenerated) in the substitute regenerated file.

106. In summary, Glasswall submits that claim 1 requires that a full parse-regenerate process is conducted in relation to each part of every incoming email: the only exception to this is that non-conforming parts may be included by operation of the threat filter. This, Glasswall said, has the benefit of meaning malformed data (including viruses and other malware) cannot be included in the new substitute file either because it cannot be parsed, or it does not conform with the rules for the relevant file type, and so cannot be regenerated. Glasswall submitted the



following diagram as a schematic of what the Patent teaches:

Findings on construction of the Patent

107. I agree with Clearswift that the concept behind the Patent is a simple one, but it is not, in my judgment, the construction for which Clearswift contends. I prefer both Glasswall's construction of claim 1 of the Patent, and its summary of the inventive concept. Both are more consistent with the words of claim 1, with a purposive construction of those words, and with the expert evidence before me.

108. In my judgment, Clearswift's contentions stretch the words of the claim beyond their natural meaning (what Glasswall's counsel described as an "unacceptable gloss"), and ignore important language in the claim. For example, the claim specifically mentions "regenerating the conforming parts of parsed content data" – this is a core concept which cannot be ignored, and is not adequately summarised in the construction for which Clearswift contends – indeed, it is ignored. The skilled person would, in my judgment, understand "regenerate" to mean more than simply "any way in which the now-approved part could be incorporated in the output".

109. On the other hand, Glasswall's construction better reflects what the skilled person would be taught by the Patent. Mr Shipp accepted the various aspects of the teaching of the Patent (taken from both the specification and the claim) when they were put to him in cross-examination:

    (a) The method functions by taking apart and putting back together again (parsing and regeneration);

    (b) The output file is a substitute file, not the file itself;

    (c) There is a change from an unknown quantity to a known quantity; and

    (d) There is a benefit to taking the file apart and putting it back together again (whilst noting the difficulties of it). Mr Shipp accepted in cross-examination that parsing and regeneration has a benefit – it "probably would remove a lot of nasties".

110. Mr Shipp also accepted that the diagram at paragraph 107 accurately reflects what is "going on" in the Patent, including the notion of parsing and regeneration. I do not accept Clearswift's counsel's submission that this meaning of parsing and regeneration is implied from the most sophisticated meaning in the specification into the claim. To the contrary, I consider the meanings submitted by Glasswall to be clear from the context. They certainly align with a purposive construction.

111. It follows that I do not accept Clearswift's construction that the subject matter of the claim is the "non-correcting" system referred to above. As averred by Clearswift, its "non-correcting" system uses only a single rule set to form a "conformant parser" that can only check that each part fully conforms with the defined rule set. This can be used with a regenerator that only amends parts to do the necessary updating (of headers and the like) when non-conforming parts are omitted. Clearswift's counsel submitted that this can be found in [0013] of the Patent, and had also been established with Professor Mitchell in cross-examination. [0013] of the Patent reads:

> "According to one aspect of the present invention, there is provided a method of receiving an electronic file containing content data in a predetermined data format, the method comprising the steps of: receiving the electronic file, determining the data format, parsing the content data, to determine whether it conforms to the predetermined data format, and if the content data does conform to the predetermined data format, regenerating the parsed data to create a regenerated electronic file in the data format."

112. I do not read in that excerpt the "non-correcting" system for which Clearswift's counsel contended. Neither did Professor Mitchell. Having read out [0013] from the Patent, Clearswift's counsel then asked "What I suggest to you, professor, is that here, rather than a

two-stage process with a reader with one set of rules and a writer with a different set of rules, we have a single stage of parsing the content data to determine whether it conforms to the predetermined data format, and if it does, proceeding then with the regeneration, That is right, is it not?" Professor Mitchell answered "I am afraid I still see two-stages here, one of which is parsing and one of which is regenerating. That roughly corresponds – more than roughly, that corresponds to the two steps of reading and writing."

113. Glasswall's counsel submitted that Clearswift could not rely on this answer and the paragraphs of cross-examination that follow for the proposition for which Clearswift contents, and my own re-reading of the transcript concurs with Glasswall's position. I do not understand Professor Mitchell to have agreed to the "non-correcting" system for which Clearswift contended. Glasswall's counsel also submitted that the non-correcting system is not apparent from Mr Shipp's three witness statements. In my judgment, the support Clearswift advanced for its contention has not been made good. I reject it.

114. Further, I do not consider that there is anything in Counsel for Clearswift's submissions that the Patent's method does not deal specifically with attachments. He suggested it would be impossible to deal with all types of attachments, and that must be right. But the clear wording of the claim (for example the reference to "a list of file types" in integer 1m) make it clear, in my judgment that the Patent teaches the checking of attachments.

115. I must also say something brief about the positioning of the threat filter. Clearswift submitted that the claim could not be read as a strict "recipe" and hence there was nothing to support Glasswall's submission that the filter is applied only after the earlier step of parsing but before regeneration. In summary, Glasswall's submission was that an attempt would first be made to parse the parts of the file – even those whitelisted on the threat filter. Professor Mitchell explained that this had advantages in catching some exploits. I accept Professor Mitchell's evidence. Mr Shipp accepted this in a limited degree in relation at least to Word files. I therefore accept Glaswall's contention that the claim requires the threat filter to be applied only after the step of parsing, and before regeneration, and I reject Clearswift's submissions to the contrary.

116. I therefore accept Glasswall's proffered construction and inventive concept.

Pozzoli's third stage

117. The parties were agreed that, in following the *Pozzoli* stages, having construed the Patent, and having articulated the relevant inventive concept, I should have regard to that inventive concept in my analysis of obviousness over the prior art. I propose to adopt that course, mindful not to lose sight of the full construction of claim 1.

118. The third stage of *Pozzoli* is to "[i]dentify what, if any, differences exist between the matter cited as forming part of 'the state of the art' and the inventive concept of the claim or the claim as construed". Clearswift's counsel submitted that this requires the Court to identify what the prior art is teaching the skilled person, and then to identify the gap between the teaching and the inventive concept. I was reminded to read and understand the prior art as a person skilled in the art – not as "a lawyer equipped with a raft of authorities on the difference between express disclosure, implied disclosure and obviousness". It was emphasised that there is a need to ask "what is this prior art actually teaching the skilled person" or "what would the skilled person take away from reading this document?" I accept those submissions as to the approach I should adopt.

119. There is a risk at this third *Pozzoli* stage that either too much will be read into the prior art, making fourth stage obviousness more likely; or that too little will be read in, leaving an Herculean task for the skilled person to bridge the gap with the inventive concept without a scintilla of invention. Counsel for Glasswall put this another way: he emphasised the importance of not conflating the issues of disclosure (stage 3 of *Pozzoli*) and obviousness (stage 4). He submitted that when identifying the differences between the prior art and the

inventive concept, it is necessary to identify what the particular piece of art relied on clearly and unmistakably discloses, in a *General Tire* sense. He referred me to Lord Walker's speech in *Synthon BV v SmithKline Beecham plc* [2006] RPC 10 at paragraph 63:

> "What emerges from the authorities, to my mind, is that enabling disclosure is a compendious summary of two distinct statutory requirements, which arise (as a pair) in two different statutory contexts: explicitly in section 14 (requirements for a patent application) and implicitly (as decided by the Court of Appeal in *General Tire & Rubber Co v Firestone Tyre & Rubber Co Ltd* [1972] RPC 457 and by this House in *Asahi*) in determining the state of the art, whether for the purposes of anticipation (section 2(2) and (3)) or obviousness (section 2 as restricted by section 3). This produces a degree of symmetry in the law and avoids divergence from the practice of the European Patent Office."

I have been mindful of his Lordship's speech in coming to my findings as set out below.

120. Counsel for Glasswall also cautioned against relying on implicit disclosures, saying that the differences must first be identified, before it is asked whether they are obvious. In the end, in this case, in my judgment it does not matter. Counsel for Clearswift submitted that I do not need to grapple with the thorny issue of where implicit disclosure ends and obviousness begins, and I agree.

Pozzoli's fourth stage

121. Stage four of Pozzoli asks the question for which the first three stages prepare the tribunal:

> "viewed without any knowledge of the alleged invention as claimed, do those differences constitute steps which would have been obvious to the person skilled in the art or do they require any degree of invention."

122. The person skilled in the art, as agreed by the parties, is set out above.

123. There was no disagreement between the parties as to the law on obviousness – I was referred to much of chapter 12 of Terrell on the Law of Patents (18th edition, Sweet & Maxwell, 2016) in which the learned authors cite from the key English cases in this area. I do not repeat those passages here, but I have taken them into account in my assessment of obviousness. As Pumfrey J set out in Glaxo Group Ltd's Patent [2004] RPC 43 at para 41:

> "It is a question of fact in every case. Both the Scylla of considering nothing obvious except that to which the skilled man is driven and the Charybdis of considering every invention obvious that can be decomposed into a sequence of obvious steps must be avoided. The former is unfair to industry because it stifles natural development. The latter is unfair to inventors and not countenanced by English patent law."

Cohen

124. Cohen is a United States patent for "preventing exploiting an email message." Like the Patent, Cohen is concerned with preventing the spread of unwanted code. The abstract of Cohen reads as follows:

> "**Method and System for preventing exploiting an email message**
>
> The present invention relates to a method for preventing exploiting an email message and a system thereof. The method comprising: decomposing the email message to its components; for each of the components, correcting the structural form (e.g. structure, format, and content) of the component to comply with common rules thereof whenever the structural form of the component deviates from the rules; and recomposing the email message from its components (in their recent state). The rules relate to email messages structure, for preventing malformed structure of email messages, for preventing exploiting an email

> message, etc. In case where the structural form of the component cannot be identified, the component may not be included within the recomposed email message, or included as is to the recomposed email message."

What does Cohen disclose to the skilled person?

125. I have above accepted Glasswall's construction of the Patent. I therefore need to compare that construction with what is disclosed clearly and unmistakeably in Cohen.

126. Clearswift's position was that Cohen teaches exactly what it (Clearswift) says the Patent teaches (in so far as it was not already part of the common general knowledge) – first, that one of the ways that malware gets control of downstream software is by exploiting poorly written parsers that respond inappropriately to unusually formatted files, and then, second, preventing unusually formatted files reaching downstream software can reduce the risk of vulnerabilities in downstream software being exploited. At the hearing I raised with Counsel for Clearswift my concern about the dangers of defining the fundamental concept at too high a level. The summary of Clearswift's position set out in this paragraph demonstrates exactly that danger.

127. For Clearswift, the only gap to be bridged between Cohen and the Patent is the threat filter. Clearswift submitted that where Cohen says a component may be "included as is" that anticipates the threat filter in the Patent. Clearswift submitted that Cohen anticipates the Patent in all respects except the *specific implementation* of the threat filter. Clearswift's counsel submitted "[s]imply making the decision identified in Cohen pre-programmable (in a known way) cannot be an invention".

128. Glasswall concedes that Cohen discloses a method of dissecting and reassembling email messages so as to correct their structure, format and content. Glasswall submitted that Cohen works as follows:

    (a) the email is split into its components and the system deals with them individually until they have all been processed;

    (b) a decision is taken whether the component complies with rules concerning structure and content;

    (c) if it does, the component is kept for later re-construction into a composed email. If it does not, it is worked on so as to comply with the structural/content rules;

    (d) if the component cannot be corrected it is blocked;

    (e) whilst this process is being carried out, the component can also be tested for hostile content; and

    (f) once every part has been checked, the process terminates.

    Further, Glasswall said that Cohen does not disclose:

    (a) Any analysis of email attachments;

    (b) Determining a "purported predetermined file type": Glasswall submitted that Cohen only deals with emails and so only has one type of parsing – a check for RFC/MIME compliance;

    (c) Parsing the component to get to its "content" (strictly so called);

    (d) Regeneration of components. Glasswall accepted that Cohen teaches regenerating emails from components, but not a process of regenerating the components themselves. As set out above, Glasswall submits, and I accept, that the Patent requires both forms of regeneration; and

    (e) A threat filter.

129. I agree. Cohen deals only with emails: it does not teach in relation to attachments: I do not accept Clearswift's submission that Base 64 encoding is an attachment. Cohen teaches some parsing, but not substantive parsing, and not regeneration. It does not teach a substitute regenerated electronic file. Cohen also does not teach that there is a benefit *per se* to parsing and regeneration. And it does not teach a threat filter of any complexity (if at all). I do not accept that the skilled person would clearly and unmistakably understand the words "included as is" to teach a whitelisting threat filter - all it says is that some non-conforming content gets passed through regardless. Cohen does not teach how that determination is made: it does not teach a dual file type/sender whitelist to be applied after parsing and regeneration have been attempted.

Is the Patent obvious over Cohen?

130. As set out above, I have accepted Glasswall's construction of the Patent. I also accept Glasswall's position on what the skilled person would take from Cohen. Can that identified and substantial gap be bridged without a scintilla of invention? In my judgment, it cannot.

131. I have found above that the claim covers emails with attachments, and Cohen does not. Clearswift's counsel submitted that no invention would be required for the skilled person to appreciate that the generality of Cohen's teaching could be extended to attachments. He was supported in this by the evidence of Mr Shipp, but not that of Professor Mitchell. I accept Clearswift's counsel's submission on this point – in my judgment, it would have been obvious to the skilled addressee to extend Cohen's teaching to emails with attachments. By the relevant date, email attachments were extremely common and were a known source of malware. It would have been an obvious extension of Cohen to apply its teachings to attachments.

132. I have also found that Cohen does not disclose determining a predetermined file type – Cohen only teaches checking for RFC/MIME compliance. If it would have been obvious to the skilled addressee to extend Cohen's teachings to emails with attachments, it would also have been obvious, in my judgment, to determine the file type of those attachments (to the extent possible).

133. Where, in my judgment, Clearswift's obviousness case falls down is in relation to parsing and regeneration, and the threat filter. As Professor Mitchell stated in relation to Cohen in cross-examination "it is not clear that there is any indication that things are ever read in accordance with a set of rules". I have found that parsing and regeneration are key elements of the Patent, and that they are not found in Cohen. This is a significant difference that would not have been obvious to the skilled addressee. As Mr Shipp accepted, the parsing/regeneration process has advantages – advantages that the skilled person does not learn from Cohen, nor which the skilled person can develop from Cohen without a scintilla of invention.

134. Turning now to the threat filter. The Patent operates on a whitelist of both sender and file type, which operates only after the file has been parsed and found to be non-compliant. Clearswift conceded that the closest Cohen gets to this is to say that non-conforming components may be "included as is to the recomposed email message". Clearswift submitted that bypassing would be immediately apparent to the skilled person who was, as I have found, familiar with whitelist bypassing so as not to block either known files or files from a known sender.

135. In my judgment, it would have been obvious to the skilled addressee to combine the two whitelists by file type *and* by sender. Although not part of the common general knowledge, as accepted by Professor Mitchell, that would not have required a scintilla of invention. The difficulty for Clearswift's argument is that I have found that Cohen does not contemplate working on different file types – hence a whitelist that allows through all RFC/MIME emails from known senders would be of limited utility, as all emails are in an RFC/MIME format using ASCII encoding. Further, I do not accept that it would have been obvious to the skilled addressee to apply a bypass whitelist only to files which had otherwise failed the

parsing/regeneration process. Clearswift submitted that some level of parsing is required to determine what the file type is and who the sender is. That much is clear. But file type and sender can readily be determined without a full parse to the content level and certainly without any regeneration. As Mr Shipp conceded, there are advantages to parsing and regenerating files that can be parsed and regenerated. Using the whitelist to let through only files that failed that process was inventive: it has distinct benefits over a whitelisting system that simply lets through all known files from a known sender.

136. In my judgment, the Patent is not invalid for obviousness over Cohen.

Avecho

137. Avecho is a series of posts on an Internet bulletin board. There are seven posts within the series, posted over a number of days, but both parties focussed their submission on the seventh post. As noted above, the seventh post was written by Mr Shipp. It is common ground that Mr Shipp had been provided with limited duration trial access to the then available Avecho anti-malware software, and had been able to test it by sending and receiving emails. His post is therefore a summary of his findings following that testing: indeed, the post itself acknowledges as much with the statement "what follows are my deductions from the emails I sent through, and therefore may not correctly reflect the actual behaviour of the system". As counsel for Glasswall succinctly put it: "most of the time Post 7 is not even reaching conclusions. It is saying 'When I tried to do this, this is what happened'. Is does not go further than saying, 'Therefore, what the engine must have been doing behind the scenes is something else'." Mr Shipp agreed with that statement in cross-examination.

138. Avecho also includes excerpts from the marketing materials related to the software. So Avecho is a mix of third party comments based on using the software, and the software producer's advertising about what the software can do. It was agreed that the discussion would be of particular interest to the skilled person, including because of the claims made for the software, and the reputation of the people who wrote the posts.

139. Both parties submitted, and I accept, that all the posts should be read as a whole.

140. Glasswall's counsel submitted that it is not appropriate under the third stage of *Pozzoli* to consider how the product might have functioned or of obvious ways of achieving the observed functionality – rather, the appropriate task for the tribunal is to determine what is disclosed in relation to the underlying method on the basis of the various observations as to the behaviour of the system on the different file types tested. There is a gap that needs to be filled between what Avecho explicitly discloses (the effect Mr Shipp saw and noted in his post) and what Clearswift says is implicitly disclosed (the method by which the effect is achieved). That gap can only be filled by the common general knowledge, or, if the skilled addressee is pointed in a particular direction, reliance can also be placed on information that would be acquired "as a matter of routine": *KCI Licensing Inc v Smith & Nephew plc* [2010] EWHC 1487 (Pat). I accept these submissions.

What does Avecho disclose to the skilled person?

141. It was common ground that Avecho disclosed that the software in question conducted some form of filtering and alteration/re-writing of emails, and that it also was able to recognise different file types and process them accordingly.

142. Clearswift submitted that Avecho described a system that receives incoming emails containing a variety of content (including email attachments). The system:

    (a) permits some content to pass through;

    (b) amends other types of content before passing it through; and

    (c) blocks other types of content completely.

143. Clearswift set out its position in detail:

> "The empirical operation of the system is described mainly in paragraph C of Post 7, in particular:
>
> a. an attachment is blocked in its entirety if it is of an unrecognised or encrypted file type, or if it is of a recognised file type which contains exclusively executable content (e.g. a file in EXE format);
>
> b. text files are recognised and are blocked if, for example, they contain the 0x7F (or 'delete') ASCII control character, or if there are not enough spaces in the text;
> c. if only part of an attachment contains executable content (an "Office document with macros" is given as an example), the non-executable part is allowed to pass through whilst the executable part of the attachment (that is, the macro) is blocked;
>
> d. attachments that contain spurious or malformed data can be rewritten so as to remove and/or correct that data, for example, spurious data at the end of BMP picture files is removed, and whitespace at the end of each line in TXT text documents is removed;
>
> e. certain 'nasty' HTML tags are stripped from HTML emails, and if an email contains HTML script then the entire email is blocked;
>
> f. malformed email headers are corrected; and
>
> g. it does not reply on signature dependency, or in fact "any form of virus detection at all.""

144. Clearswift submitted that the skilled person would know from this description that the software must parse each file by reference to the rules applicable to the file type. The skilled person would understand that in addition to checking against the file specification the system would also apply further limitations (such as stripping "nasty" HTML tags). Further, Clearswift says that the skilled addressee would understand the large number of false positives thereby generated – hence the need for the "trust bypass" feature.

145. Clearswift's case was that Avecho discloses to the skilled person the whole of claim 1 of the Patent.

146. Glasswall conceded that Avecho discloses that the engine is able to deal with different file types: various file types are mentioned in Post 7. Post 7 then goes on to discuss the different behaviours of the engine on different file types, for example: EXE files are blocked; Office documents have their macros stripped; "nasty" tags are stripped from HTML files; some rewriting occurs on BMP and TXT files; RAR files are blocked; and encrypted files are blocked. From this, Glasswall accepts that Avecho discloses that the software conducts filtering/re-writing of emails, and does so by recognising different file types and processing them accordingly. But, beyond that, Glasswall's counsel submitted that very little is clear. Glasswall considers that integers 1d, 1e, 1f, 1g, 1h, 1i, lj, lk or 1l are not disclosed.

147. I agree with Glasswall that it is difficult for the skilled addressee to extract much from Avecho that is clear and unmistakable. I accept that the skilled addressee would have been very interested in the posts, but the skilled addressee would, in my judgment, not have taken from it what Clearswift submits. Specifically, in my judgment, Avecho does not disclose a number of teachings of the Patent.

    (a) Whilst the software would assess file type for the purpose of scanning and removing unwanted code, this would not be sufficient to parse the file so as to extract the content data;

    (b) Avecho does not disclose parsing of each part in accordance with the relevant specification followed by regeneration;

    (c) Avecho does not disclose parsing to content level. Whilst Avecho discloses stripping macros from Office documents, stripping "nasty" tags from HTML

emails etc, none of these requires a full parse of the file, or a parse to get down to the content level;

(d) Rewriting is not the same as regeneration: Avecho does not provide for regeneration; and

(e) Whilst Avecho discloses a threat filter, it provides none of the detail as to the threat filter of the Patent as set out above.

Is the Patent obvious over Avecho?

148. I have found, contrary to Clearswift's submission, that Avecho discloses comparatively little. For the same reasons as set out above in relation to parsing/regeneration and the threat filter in Cohen, I do not consider that the additional disclosures I have found the Patent to make would be obvious over Avecho.

Conclusion

149. In my judgment, for the reasons set out above, the Patent is not obvious over Cohen or Avecho. Clearswift's application for revocation of the Patent is therefore dismissed.

Post Script

150. Following my provision of a draft copy of this judgment to the parties in the usual way, Clearswift's counsel provided me with a 38 paragraph document expressly said to be pursuant to his duty under *Re M* [2008] EWCA Civ 1261 "to raise with the judge not just any alleged deficiency in the judge's reasoning process but any genuine query or ambiguity which arises on the judgment. Judges should welcome this process." To address the alleged deficiencies of reasoning, I added a number of paragraphs to this already overly-long judgment to attempt to deal with counsel's concerns.