Electronic Watermarks to Help Authenticate Soft-copy Exams

Simon
School of Design, Communication, and Information Technology
University of Newcastle
PO Box 127, Ourimbah 2258, New South Wales, Australia
simon@newcastle.edu.au

Abstract

When conducting remote web-delivered exams in a distance learning course, we make the exams available over a period of 48 hours, allowing the students to choose when in that period they will undertake the exam.

This paper describes an electronic 'watermarking' system that we have used to check whether the completed exam returned by a student is the same exam that the student downloaded, or whether it is one that was downloaded earlier by another student.

Findings from the use of the system lead to some disturbing questions about the practice of distance education and education in general.

Keywords: assessment, authentication, distance learning.

1 Introduction

Some of our web-based distance courses include web-based distance exams, typically of 2 or 3 hours duration. To accommodate different student requirements, we make the exam available over a period of 48 hours. We are able to determine when the student downloads the exam and when the student returns the exam, and can thus check that the student completes the exam within the specified time limit.

Of course, the 48-hour period of availability means that some students will consider downloading and completing the exam early in the period, then passing it on to fellow students, giving them a chance to prepare it before downloading their own copy.

We therefore provide multiple versions of the exam, effectively indistinguishable to the students. When a student returns an exam, we check whether it is the same version that the student downloaded. If it turns out to be an earlier version, we know that the student has cheated, and take appropriate action.

In this paper, the word 'course' is used to mean a single unit of study that leads to a formal result. Such courses are also sometimes known as subjects or papers.

In places, this paper is a highly personal account. At these places, it is written in the first person singular ('I') rather than the traditional academic first person plural ('we').

2 Why do we set exams?

In many university courses, the final examination is the ultimate assessment of the student's knowledge and skills pertaining to the subject matter. It is the means for students to show what they have learnt, for academics to assess what the students have learnt (Thorpe, 1998), and perhaps even for some cynical academics to show the students how little they have learnt.

But many computing courses have a strongly practical orientation, which is better assessed in assignments (formative assessment[1]) than in exams (summative assessment). Why conduct exams in these courses?

A common answer (eg Thomas, 2001) is that while we can never be sure just who has written a student's assignments, we can almost always be confident that a student has written his or her own exams. For this reason, even in highly practical courses such as programming or database implementation, we tend to set exams and to allocate a high proportion of the final mark to them.

This confidence is occasionally misplaced, and one hears reasonably often of students paying other people to take their exams (Zobel, 2004; Fröhlich, 2000); but the general principle still applies, and we set exams to improve the likelihood that we are assessing the work of the right person.

The value of this approach is still under discussion. Some authors suggest that 'current assessment practices in higher education are long overdue for a re-think; they are particularly ill-suited to the digital age' (Mason, 1989, as cited in Ruhe, 2002) and that success depends on 'a progressive shift from summative to formative approaches and reaching a balance of both' (Lim, 2004). The opposing viewpoint is that 'a distance learning programme must be based on an existing [face-to-face] programme, which is the gold standard for curriculum and its assessment, and where possible, common examinations and assessments [should] be used.' (Clarke, 2004).

[1] In the terminology of education, assignments generally fall into the category of 'formative' assessment, which is used substantially to help students develop their skills or knowledge, whereas exams are generally 'summative' assessment, which is used principally to rate and/or rank students.

Notwithstanding the debate, many academics continue to use exams in the belief that they are more readily authenticated than assignments.

3 Techniques for conducting exams in remote online courses

Thomas (2002) observes that 'a common criticism of remote examinations used for summative purposes is the difficulty of ensuring that cheating is minimised'. Universities have developed a number of techniques for conducting distance exams with this in mind. Some of them are surveyed briefly in the remainder of this section.

3.1 The cooperating institution

It is not uncommon for a university to ask cooperating institutions in the students' neighbourhoods to invigilate exams.

This technique offers about the same level of security as an exam invigilated at the university itself. Unfortunately, it takes time and effort to establish a network of cooperating institutions. Thomas (1998) observes the difficulty faced by the Open University in making special arrangements for students outside the UK.

Furthermore, the advent of web-based online courses has spread the student body so wide, and yet so sparse, that the establishment and maintenance of a usable network of institutions would be all but impossible.

By way of illustration, in one of our courses last year it would have been easy to arrange invigilation for the 15 students in five different locations within Australia, and reasonably easy for the dozen in Singapore and the half dozen in Hong Kong. But it would have been somewhat more difficult to find cooperating institutions for the students in Botswana, Central Africa, China, Malaysia, Netherlands, Norway, Papua Niugini, and the United Arab Emirates.

3.2 The trustworthy third party

A number of universities ask the students to arrange a trustworthy third party to invigilate the exam. In one arrangement used at our university, the invigilators have to sign statutory declarations attesting to their trustworthiness.

Am I the only one who sees a certain flaw in this system? The statutory declaration could equally well indicate a trustworthy person asserting that trustworthiness or an untrustworthy person who has no qualms about lying.

3.3 The webcam

A researcher at the University of Newcastle, Australia, has developed a system whereby the invigilation is performed remotely with the help of a webcam. No reference is available for this, as we believe it is the subject of a patent application rather than an academic paper.

While students are doing the exam, the webcam is turned on and transmitting pictures of them to an invigilator back at the University.

Am I the only one who sees a certain flaw in this system? If the student writes a note and moves it to one side, out of the camera's field of vision, does the invigilator know whether this is work in progress or communication with another person? Can the invigilator see the student's screen, to know whether the student is doing the exam, browsing the web for possible answers, or emailing a friend?

Fröhlich (2000) reports on a number of technological systems, including a patented 360° webcam system and other options such as biometric recognition, fingerprint recognition, retinal scans, and others. I recognise the possible value in all of these, but immediately set about imagining ways in which they might be subverted.

3.4 Trust

Some Universities trust the students to do the exam honestly and fairly.

Am I the only one who sees a certain flaw in this system?

Some academics start with a basis of trust, but supplement that so that students are aware that there will be some surveillance. Thomas, Price, Petre, Carswell, and Richards (2001) suggest telephoning a selection of students immediately after the exam and asking them questions relating to the exam or to their own answers. They suggest that this would amount to 'an honour system in which the student was aware that measures would be taken to detect cheating and that there was a reasonable chance of being caught out.'

4 Constraints on exams in remote online courses

It is accepted that in face-to-face courses, the examination will be held at a specified time and place, and all students will attend. Other arrangements will generally be made for students prevented from attending by dramatic and unforeseen circumstances, and even sometimes for students prevented from attending by foreseen circumstances, but for the bulk of the students the exam is at a fixed time and place.

One of the advantages of an online course is that students can 'attend' at times that suit them. Indeed, it is for this reason that students will sometimes prefer an online course to a face-to-face one offered by an institution in the same city they live in. It is a reasonable expectation that this flexibility of attendance will extend to an examination if there is one.

Further, in a given online course the students will be spread across different time zones. The class described briefly in 3.1 took in six different time zones with a span of 10 hours. Further still, some of the students will be studying from home while others study at work.

For these reasons it is generally impractical to conduct the exam at a single time (what Thomas (2002) calls a 'synchronous examination'), and we must resort to an

'asynchronous examination', in which students choose their own sitting time from a specified range.

5 Conducting a timed remote online exam

In the courses that this paper deals with, an asynchronous timed exam is run in the following manner.

The exam is available for 48 hours, covering the best part of a Sunday and a Monday in most time zones, to cater for most student timing requirements.

Students are asked to download the exam some time during the 48 hours, to complete it by writing their answers in the same document that was downloaded, and to return it. They are given a time limit that includes the notional exam time and an additional 15 minutes or so to allow for internet connections, downloading, and uploading. Exceeding the time limit is penalised, typically at 2% for every minute by which the limit is exceeded.

When the 48-hour period begins, students are given a URL, which they are to access when they are ready to sit the exam.

At that URL is a password-protected form with a message, some text fields, and a button. When the button is clicked, the form is checked to see that the required fields (name, student number, and password) have been filled with plausible data, and an email is sent automatically to the lecturer, with the time and the contents of the name and student number fields. The student is then taken to a second form containing a link from which the exam document can be downloaded.

The student adds answers between the questions in the document, then returns it via the Digital Dropbox feature of Blackboard™, which puts a time stamp on its return.

The download time, from the automatic email, and the return time, from Blackboard™, are combined to establish the time taken by the student to complete the exam.

6 Some ways of cheating in a remote online exam

As in most academic situations, the number of ways of cheating is limited only by the imaginations of the students.

Students might try to bypass the form that sends the automatic email, thus accessing the exam without the lecturer's knowledge. Having prepared their answers at their leisure, they will then download the exam in the expected way, complete it, and return it in good time.

A simpler way of arranging early access to the exam is to have it sent on by a fellow student who downloaded it early in the period of availability.

Students might cheat by having other students or mentors with them when they do the exam, and taking advice or assistance from these extraneous people.

Students might cheat by emailing other students for assistance during the time they are doing the exam.

7 The electronic watermark to detect one form of cheating

Believing that early access to the exam was perhaps the likeliest form of cheating, we devised a system that might help to detect such cheating.

The principle of the system is that there are many versions of the exam paper, indistinguishable to the students, and that different versions are made available at intervals during the period of availability.

When a student's exam is returned, it is checked to see whether it is the same version that was available at the time the student downloaded the exam. If it is an earlier version, the student must have had access to the exam before downloading it in the expected manner.

It is integral to this approach that the papers must be indistinguishable to the student. With the alternative approaches of permuting the questions or choosing a different set of questions from a larger bank, and possibly even having a different filename for each version, the (hypothetical) cheating student would immediately notice the difference and act accordingly. Only if students believe the exams are identical are they likely to return the earlier version to which they had unauthorised access.

8 Development of the watermark system

A number of possible watermarks were considered and adopted. Minor variations in the text, such as a double space between two particular words, were one possibility. Another was a component of a header or footer, invisible by virtue of being in white text on a white background and by virtue of being in the body of the page rather than in the normal header and footer regions.

But while these stratagems mark the document as a whole, they do not mark all of its text. They will detect a student who accesses the exam early, completes it, and returns it after downloading a later version. But they will fail to detect a student who accesses the exam early, completes it, then copies and pastes the answers into the later version.

Our search for a definitive watermark settled on the colour of the text in the document. While text colour is generally chosen from a limited palette, in some word processors it can be specified by RGB values.

Each of our versions of the exam paper has its text in a colour that has a distinctly different RGB setting, but that to the eye is indistinguishable from black.

For example, version 1 has its text in RGB 0,0,1, version 2 in RGB 0,1,0, version 3 in RGB 0,1,1, version 4 in RGB 1,0,0 and so on. Because all of the text is in these colours, answers typed in the document by the students will have these same colours.

8.1 First pass: successes and problems

Our first use of the watermark system, in August 2003, had some impressive successes and brought to light some significant problems.

Ever optimistic, we were actually hoping not to find any cheating. Of course we are also realistic; otherwise we would not have devised the watermark system. So we were not surprised when we did find evidence of cheating.

Over the 48-hour period of availability, 15 versions of the exam were made available, each replacing the previous one at a time that was noted accurately.

It was intended to replace the exam each time a download was notified to the lecturer; but that weekend the University's email system suffered a breakdown, and the notification emails took up to 15 hours to arrive. Hence the decision to make new versions available at whim, and to record the times scrupulously.

The first anomaly was a version 1 exam returned by a student – let us call him B – who downloaded version 9. Version 1 was available from midnight until 8am on the first day; the student downloaded version 9 of the exam at 9:15am the following day; so he had at least 25 hours to do the exam instead of the specified three hours.

This case immediately brought to light the first problem with the system. Version 1 of the exam had been downloaded by two students, whom we shall call A and X. It was absolutely clear to the lecturer, for several reasons, that it was A who had provided the exam to B; but the watermark system could not provide the evidence to back this up.

In a second anomaly, student K downloaded version 13 and then returned version 11. The timings showed that he had had a little more than seven hours to do the three-hour exam.

Four students had downloaded version 11, and this time it was not at all clear which of them had provided the exam for student K.

Students B and K were given zero for the exam, the strongest penalty that the University was prepared to impose. Student B, who still passed the course, accepted the finding without argument. Student K, who failed the course, protested his innocence long and loud; but the University hierarchy was persuaded of the validity of the watermark system and stood by the finding. Student K was never informed that he would have failed the course even if his non-zero mark for the exam had been restored.

Student A would have failed the course had she been given zero for the exam. As it was, she just scraped through.

Contemplating why B, one of the stronger students, would cheat in this way, we turned back to A's exam and found that while some of the answers bore the correct watermark, others bore no watermark at all. Further, they appeared to have been answered in a different style. This suggested that these answers had been written by somebody else and copied and pasted into A's exam. We were left concluding that A and B, and possibly further students, had worked together on A's exam; and that B had then composed his own answers to the exam at his leisure before going through the motions of downloading it the next day.

A different anomaly that arose was that some of the exams came back in colours that were clearly nowhere near black, having been turned by some software into colours from a standard palette. For example, two version 4 exams (RGB 1,0,0) were downloaded; both came back as RGB 128,0,0; of the six version 13 exams (RGB 3,0,1) that were downloaded, one came back as RGB 128,0,128.

It is now clear to us that the software that is doing this transforms any small non-zero R or G or B value to 128. This was not clear at the time because the first one we received was downloaded as version 2 (RGB 0,1,0) and returned as RGB 0,0,128.

Only when we belatedly discovered the system behind the colour changes did we realise that this was 'the third man', student C, a further collaborator with A and B on A's exam. This discovery came too late to penalise C, who passed the course – and would have just scraped through even if he'd been given zero for the exam.

We do wonder what student C thought when the exam that had appeared black on A's computer appeared dark blue on his, but we have never asked, for fear of alerting students to the nature of the watermark. Student K, by the way, was admirably insistent that we tell him the nature of the watermark. We are pleased to say that the University declined to acquiesce to his insistence.

8.2 Second pass: successes and problems

The clear problem with the first pass is that if several students downloaded the same version of the exam, we were able to catch the cheating student who wrongly returned that exam, but not the cheating student who provided it (and perhaps benefited from assistance while doing it).

It should clearly be possible to write a system that automatically uploads a new version of the exam every time the current version is accessed. We do not have the time to work out how to write such a system. We asked the University IT infrastructure staff to write such a system, and they were not able to do so.

So the second time we used the system, in May 2004, we were still in the position of having to keep close watch on our email over the 48-hour period of availability and upload a new version of the exam every time we were notified that the exam had been accessed.

In addition, we were given access to the University's web logs for the period, and were able to determine that one person tried unsuccessfully to bypass the authentication process and access the exam directly. Given the time of the attempted access, the location of the computer from which it was attempted (within the University), and the username under which the attempt was made, we concluded that it was an opportunistic attempt by somebody who found a computer left logged in.

One anomalous return was noted for this exam. Student M, who had downloaded version 20 of the exam, returned version 10. Unfortunately, this brought a new problem to light.

In the first pass, when several students downloaded the same version of the exam, we had confirmation that the correct version had indeed been made available at the time in question.

Now that each version was downloaded by only one student, it was possible that the lecturer had made a mistake, uploading version 10 instead of version 20.

This interpretation was supported by the lecturer's judgement that the student who downloaded version 10 would not have been involved in cheating. Student M was not penalised. He passed the course, which he would not have done if he had scored zero for the exam.

8.3 Third pass: successes and problems

The third time we used the watermark system, in August 2004, we were still stuck with the email observation and manual uploading. The University's email system was variable: most messages were coming through within a minute or two, while a handful took up to four hours.

We supplemented the processes used previously with one further task: every time a new exam was uploaded, a copy was immediately downloaded and given a name that indicated its supposed version and the download time. Any doubts as to which version had been available at a particular time could be resolved by checking these downloaded versions.

There was only one anomaly this time, one that caught us somewhat by surprise.

Student Q downloaded version 10 of the exam and returned version 10, with the exception that one part of one question bore the watermark of version 7.

The student who had downloaded version 7 did not seem a likely suspect, so we had to think a little more laterally. We checked the previous year's records and found that a plausible suspect, P, had downloaded version 7 when he did the course the previous year. It was easy to establish that the wording of Q's answer for 2003 was identical to P's for 2004. This was somewhat amusing, as the questions, while sharing a surface similarity, were actually quite different.

Student Q was not penalised for cheating, as he had not collaborated with another student in the course. Instead he was given zero for that question, as he had contravened the requirement to fully reference any use of resources other than his own recollections.

Student Q failed the course, and would have done so even without this penalty.

9 Questions arising from use of the watermark system

9.1 Is it worth the effort?

In its current form, the watermarking system requires almost constant vigilance over the 48 hours that the exam is available. This has the effect of leaving the lecturer sleep-deprived, which is not the ideal state in which to begin marking exams.

The system succeeded almost spectacularly in its first use, and has not definitively detected any cheating since then.

Has word passed around that we are indeed able to detect cheating, and has the cheating subsequently stopped?

We believe it more likely that students have some idea of what sort of cheating we can detect, and that they are managing to cheat in other ways.

If this is the case, we are going to all the effort of creating multiple version of the exams and 'baby-sitting' the email for 48 hours with little likelihood of a useful return.

If we do ever manage to write, or to have somebody else write, an application that automates the version-changing process, this will greatly reduce the effort involved in conducting the exam.

One of the referees for this paper has suggested that we might be able to arrange for the full set of exams to be made available at the same start time, with the system choosing the designated exam for each student based on the student ID provided (Anon1, 2004). We shall certainly consider that option, but at first consideration it seems that it would possibly require files with different names, which would alert the (hypothetical) cheating student; or, at best, different directory paths, which would alert the (hypothetical) cheating student who thought to look at the source of the web form from which the file is downloaded.

9.2 Is there any point in conducting exams in remote online courses?

It was argued in section 2 that the principal reason for setting exams as well as assignments is to acquire some work that we can be sure is the student's own, produced in conditions that are more or less comparable with those in which other students are working.

It is evident to us that students are willing to cheat in online exams. We set up a system to detect one form of cheating, and it succeeded immediately. We assume that other forms of cheating, not detectable by this system, are also taking place.

We are left to conclude that the reason for setting exams does not apply in online courses, as we cannot be at all sure that the work is the student's own, or that all students in a course undertake their exams in comparable conditions.

9.3 Is there any point in conducting assessment in remote online courses?

By the same token, we must wonder if there is any point in setting any assessment tasks in online courses. If even the exams are not guaranteed to be the students' own work, what hope is there for the assignments, which the students have weeks to complete?

9.4 Is there any point in conducting remote online courses?

With such a cynical outlook, we must surely conclude that there is no point in conducting remote online courses; that in doing so we are assuredly at times conferring qualifications on students who have not earned them.

But there is, of course, a pressing reason for conducting online remote courses: our university believes that they bring in lots of money. While the academics who work in an institution might well be idealistic, the institution itself must of necessity be commercial.

Ruhe (2002) points out the spectacular growth in online enrolments at institutions in the USA. This trend is not restricted to that country, and many institutions are reliant on such enrolments to help them remain properly funded. Indeed, the belief appears to be that if an institution allows itself to fall behind, it will also fall by the wayside.

It is not universally accepted that universities do actually make money through online teaching. In his essay on 'digital diploma mills' (Noble, 1997) David Noble wrote: 'Experience to date demonstrates clearly that computer-based teaching, with its limitless demands upon instructor time and vastly expanded overhead requirements – equipment, upgrades, maintenance, and technical and administrative support staff – costs more not less than traditional education, whatever the reductions in direct labor.' However, this essay was written some eight years ago, principally about online education for on-campus students; institutions today, if they are aware of it at all, presumably dismiss its arguments as either wrong or no longer valid.

9.5 Is there any point in conducting courses?

The logical extension of the previous question is whether even face-to-face courses meet the academic ideal for which they were created.

We are aware of students who have passed courses not by academic achievement but by bullying and blustering and/or by cheating.

We do like to believe that such students are in a small minority, and that the work we do really is worthwhile. But if we believe that for face-to-face students, should we then persuade ourselves that it is true for remote students as well? Should we simply stop worrying about the cheats and assess as if everyone were honest?

It would certainly be a great deal easier, except perhaps on the conscience.

10 Are we interested only in detecting cheating?

One of the referees for this paper asked some highly pertinent questions that can be paraphrased as:

- Why do you not concentrate your efforts on preventing cheating in the first place?
- Why do you not direct your effort towards student learning?
- Why do you not try to alter the way in which teaching and learning happen so that cheating is assumed not to happen, and that cooperative and ethical behaviour is a natural outcome? (Anon2, 2004)

To the first two questions, we reply: we do. The work reported in this paper is one aspect of our approach, and should not be taken as suggesting that catching and punishing cheats are our only goals.

To the third question we pose a parallel question: why do the police and government not alter the way in which society works so that crime is assumed not to happen, and that cooperative and ethical behaviour is a natural outcome?

The answer to the parallel question is probably obvious. Even if it were possible to reshape society in such a way, it would be a massive undertaking that would last hundreds of generations. But many great thinkers have concluded that it is simply not possible.

It should be pointed out that assuming that there will be crime is by no means the same as assuming that everyone is a criminal. The measures put in place to detect crime and other misdemeanours (speed cameras, red-light cameras, video surveillance, etc) are based on the assumption that *some* people will err, and the desire to detect it when this happens. If anybody finds a way to recreate society so that nobody will ever err again, the measures can be dispensed with.

Sheard, Carbone, and Dick (2002) conducted a comprehensive survey in which they found that as many as 80% of students admitted to some form of cheating. Exploring the reasons given for the cheating, they concluded that students should be assisted to 'develop strategies to manage the internal factors that lead to poor learning tendencies', and that 'it is also important for educators to address external factors . . . caused by characteristics of the learning environment they provide for students.'

We have no argument with this, but we believe it would be naïve to assume that such an approach would completely do away with cheating.

Our system has clearly indicated that cheating does take place in our courses. It is our contention that so long as cheating takes place, it is better to detect it where possible, and to deal with it appropriately, than to turn a blind eye to it.

11 Conclusion

Alert to the possibility of cheating in remote online exams, we devised a system to detect that sort of cheating, and found that it does indeed happen.

We are aware of other possible ways of cheating, ways that we have not yet worked out how to detect, so we assume that these other ways of cheating are also being employed.

We do take steps to try to reduce the motivation to cheat, but we are not confident that these steps can ever completely eliminate cheating.

We are left wondering whether to enhance our system so that it covers more and more forms of cheating, if indeed that is possible, or to give up, go with the flow, and behave as if there were no such thing as cheating.

12 References

Anon1 (2004): Blind review of this paper.

Anon2 (2004): Blind review of this paper.

Clarke, M., Butler, C., Schmidt-Hansen, P., Somerville, M. (2004): Quality assurance for distance learning: a case study at Brunel University. *British Journal of Educational Technology* **35**(1), 5-11.

Fröhlich, R., (2000): Keeping the wolves from the doors... wolves in sheep's clothing, that is. *Fourth International Computer Assisted Assessment Conference*, Loughborough, UK.

Lim, C.P., Hung, D., Wong, P., Chun, H. (2004): The pedagogical design of ICT integration in online learning: a case study. *International Journal of Instructional Media* **31**(1), 37-47.

Mason, R. (1989): A case study of the use of computer conferencing at the Open University. *Unpublished doctoral dissertation*, Open University, England.

Noble, D. (1997): Digital Diploma Mills: The Automation of Higher Education. *Essay distributed by Ontario Confederation of University Faculty Associations*.

Ruhe, V. (2002): Issues in the Validation of Assessment in Technology-Based Distance and Distributed Learning: What Can We Learn From Messick's Framework? *International Journal of Testing* **2**(2), 143-159.

Sheard, J., Carbone, A., Dick, M. (2002): Determination of Factors which Impact on IT Students' Propensity to Cheat. *Proc. Fifth Australasian Computing Education Conference*, Adelaide, Australia, 119-126, ACM Press.

Thomas, P., Carswell, L., Price, P., Petre, M. (1998): A holistic approach to supporting distance learning using the Internet: transformation, not translation. *British Journal of Educational Technology* **29**(2), 1-13.

Thomas, P., Price, B., Petre, M., Carswell, L., Richards, M. (2001): Experiments with electronic examinations over the Internet. *Fifth International Computer Assisted Assessment (CAA) Conference*, Loughborough, UK.

Thomas, P., Price, B., Paine, C., Richards, M. (2002): Remote Electronic Examinations: student experiences. *British Journal of Educational Technology* **33**(5), 537-549.

Thorpe, M. (1998): Assessment and 'third generation' distance education. *Distance Education* **19**(2), 265-286.

Zobel, J. (2004): "Uni Cheats Racket": A Case Study in Plagiarism Investigation. *Proc. Sixth Australasian Computing Education Conference*, Dunedin, New Zealand, 357-365, ACM Press.