

EVIDENCE OF MOOC STUDENTS USING MULTIPLE ACCOUNTS TO HARVEST CORRECT ANSWERS

Giora Alexandron^a, Jose A. Ruiperez-Valiente^{a,b,c}, David E. Pritchard^a

^a Massachusetts Institute of Technology, ^b Universidad Carlos III de Madrid, ^c IMDEA
Networks Institute

1. INTRODUCTION:

Massive Open Online Courses (MOOCs) give “honor certificates” of accomplishment for successful completion of their courses. In part because these certificates are given without any verification of the identity of the student, such tokens have little professional value – they would seem to have value mostly for personal reasons. However, although the value of the certificates is not clear, we have found that a substantial number of students do not obey the code of honor, but apparently are using multiple accounts to harvest the correct answers for questions that they later submit with their main account. This is enabled because most MOOCs provide an instant feedback about the correctness of a student’s answers. In this paper we present evidence for, and detailed patterns of such behavior; we then briefly discuss ways to prevent it.

To the best of our knowledge, our study is the first to report on this type of cheating in MOOCs. The topic of academic dishonesty in online learning environments was addressed by various studies, see for example [1-3], but in the context of MOOCs research, it still did not get a lot of attention. We believe that it is very important to discuss the various aspects of this sensitive issue, as if it is not handled properly, it can jeopardize the value of MOOCs as academic credentials.

2. HARVESTING PATTERNS:

Our approach was to hypothesize and operationalize behavioral patterns that are likely to be carried out by users who are using multiple accounts for harvesting solutions, and then track the appearance of these patterns within the logs. We have defined two different patterns that we believe are very likely to reveal users implementing this behavior (obviously there can be additional patterns to the ones we have defined). Observations of these patterns provide strong evidence that this behavior is actually happening in MOOCs. Both of the patterns are strongly related to the use of the same IP address in both harvester and main accounts. The patterns are:

Instant Mode. This pattern refers to a situation in which the user is harvesting the correct answer with his/her harvester account and inserting the right solution in his main account. The detection criteria for this pattern are that both accounts are using the *same* IP address for answering this *specific* question (users can have multiple IPs during the course), that both have the right answer (for the harvester, possibly after few erroneous trials), and that the time between both actions is short (we used a threshold of fifteen minutes) with the harvester first. One way to implement this behavior is by using two browsers simultaneously, each one connected with a different user (in edX it is not possible to have two users connected simultaneously on the same browser).

Batch Mode. This pattern refers to a situation in which the harvester account is used to obtain the correct solution for multiple questions, and later inserting the correct solution for several of these questions in a row in the main account. The criteria we use to identify this behavior is a user (main account) who inserts a very fast sequence of (only) correct answers to questions that were previously answered by a second user (harvester account) who *shared* IP address with the first user at *some point* in the course. The threshold that we use is ten questions, with less than 20 seconds between consecutive submissions. Though we do not have a reference model, we believe that this threshold is quite strict. This pattern can expose, for example, users who are using the harvester account at one place (e.g., home), and then the main account on another place (e.g., work). Referring to an unreasonably fast submission as an indication of cheating was also done in [2].

3. PRELIMINARY RESULTS:

We implemented algorithms that detect these patterns, and applied them to the data obtained from the 2014 instance of the introductory physics MOOC 8.MReV given by the third author and his team through the edX platform. The course lasted for 14 weeks and attracted about 13500 students, from which 502 earned a certificate. The results are presented in the next table:

	Number of unique students	# questions solved this way
Instant pattern	306 students (98 certificate earners)	27297
Batch pattern	112 (57 certificate earners)	3541
Either of both patterns	353 (120 certificate earners)	30838

Our first application of these algorithms indicated that around 24% of the certificate earners used one of these two patterns at least once during the course. As we wanted to be conservative and reduce the false positive rate, we applied an additional filter that removed harvester-main couples for which either the main account also demonstrated a harvester behavior, or that the harvester was used as a main account, at some point in the course. The rationale is that such a behavior probably reveals two real users working in close proximity (e.g., dormitory) and sharing answers, which is a behavior that was not in the scope of this study. With this filter, we get that 135 main accounts used a harvester account at least once, were among them, **49 were certificate earners (9.8%)**. Among the certified students who have used this method, the average number of questions solved this way was 70, and in the non-certified group, it was 48. However the difference between the groups is not statistically significant (p-value = 0.15). Further understanding the characteristics of these two groups is one of the things that we are studying. For example, we see that some of the students are using several harvesting accounts. Also, we suspect that some students who have used this method and did not earned a certificate actually started the course with the intention of earning one, but for some reason dropped-out on the way.

4. DISCUSSION AND FUTURE WORK:

We believe that our results give a good estimation of the number of students who have used a harvester account in our course. The numbers are quite high, especially when looking on the number within the group of the certificate earners. If these numbers represent, even roughly, the percentage of students committing this behavior in other MOOCs, then this is a severe issue that calls for discussion and action. If not addressed, this issue has the potential to jeopardize the value of even verified MOOC certificates. Future issues that we intend to study involve learning more about the users who perform these behaviors, and the impact on their learning – for example, do they learn less and show lower achievements in the long run? We are also interested to study why and when students take this approach (high-stakes questions? more difficult ones? tight deadline?). We want to explore also results in the case that harvester and main account get different problems for the same item. This can be done by randomization of the problems or by using A/B testing functionality. It would be expected for users committing this behavior to perform worse on these problems than those that remain the same for both harvester and main account.

Right now the detection mode is strictly related to the use of the same IP in both accounts. One challenge would be to be able to design an algorithm that can detect this behavior even when separate IPs are used for harvesting and answering. Clearly the collaboration of two different users using the same IP address bears further investigation. Equally important, we are interested in ways for preventing this behavior. One direction is the use of randomized questions for the students, but this will put a heavier load of work on course developers, and it also requires developing scalable methods for ensuring that different students receive questions with the same level of difficulty. Another option is to remove the feedback about correctness that students receive after submitting an answer until the deadline has passed, but pedagogic-wise, this is undesirable. Ultimately, an intervention system could be applied, and this is also a direction that we believe is worth studying.

ACKNOWLEDGEMENTS:

We would like to thank Isaac Chuang, Zhongzhou Chen and Christopher Chudzicki for their useful comments.

REFERENCES:

- [1] Grijalva, T. C., Nowell, C., Kerkvliet, J. Academic Honesty and Online Courses. College Student Journal, vol 40, 2006.
- [2] Palazzo, D. J., Lee, Y., Warnakulasooriya, R., Pritchard, D. E. Patterns, correlates, and reduction of homework copying. American Physical Society, vol 6, 2010
- [3] Watson, G., Sottile, J. Cheating in the Digital Age: Do Students Cheat More in Online Courses? Online Journal of Distance Learning Administration, v13 n1 Spr 2010