

Applying Rating Systems to Challenge Based Cybersecurity Education

Andrew Samuels
School of Computing
Clemson University
absamue@g.clemson.edu

Feng Li
Computer and Information Science
IUPUI
fengli@iupui.edu

Connie Justice
Computer and Information Science
IUPUI
cjustice@iupui.edu

Abstract—As technology becomes a larger part of everyday life, it becomes increasingly more important for CS and CIT students to learn about cyber security during their education. While many cyber security oriented courses exist, it is also necessary that students must be able to work and learn in an environment that resembles a real world context. To address this problem it has become common to adapt cyber security challenges into the classroom as a method for students to put their knowledge into practice. One problem is that these challenges can vary considerably in levels of difficulty, which makes it problematic for students to be able to select a challenge that is an appropriate difficulty for their skill level. A potential solution to this problem could be to adapt a rating system to rank both the students and the challenges. This would then allow the students to easily select challenges that are appropriate for them to engage with by comparing their own rating with the rating of available challenges. In this project we propose methods that could be used to adapt a rating system to an existing cyber security education program. Finally we propose a method to survey students that interact with the program so that the effect of the rating system can be measured.

1. Introduction

Cyber security challenge competitions and events are commonly held where students or professionals can put their knowledge into practice. These competitions offer participants with a way to gauge their own knowledge against a series of set challenges, or against a group of opponents in a team based competition. While these competitions can be for fun or self-improvement, they can also be used in a classroom environment as a method to teach students. Rather than competition goals, the challenges can be made as assignments for students to work on as they learn about the related material

in a cyber security class. This then gives students the ability to use their knowledge in a realistic environment to further solidify their understanding of the material.

In many implementations, students would interact with such a program in a typical lab activity format where the challenges move along with the class material. Unfortunately this may leave students struggling in instances where they do not have a strong understanding of the underlying material, which hinders the student from learning properly. In order to combat this, we believe that a more adaptive approach could be more beneficial to the students. Such a system would be able to accommodate for the skill level of the students involved, allowing the students to work on material that will give them the appropriate difficulty challenge to work on. This will be accomplished by the utilization of a rating system that will rate both the students and the challenges, which we have proposed methods for creating and maintaining such a system.

2. JagWarz as a Base

This project is based around the use of cyber security challenge games that are designed for the classroom, such as the JagWarz program created at IUPUI. The JagWarz program is designed to provide students with a selection of challenges that vary both in content in difficulty. Students that use the program are then able to practice in areas that they are weak in, while also improving on their strengths. This variation in challenge type and difficulty is a good platform for adapting a rating system, as this provides a basis to have challenges that will have a large variation in rated difficulty.

3. Student Skill Level

One of the large difficulties when facing cyber security education with challenge based programs is the fact that not all students will be at the same skill level. Some students might only have a basic knowledge of the material covered in the program while others might have undergone enough self study to have a solid foundation of necessary skills. For the first case we do not want an inexperienced student to feel that they are not capable of participating in a challenge program. It is also undesirable for the more experienced students to feel that they have nothing to learn from the program. Consider the situation in which available challenges are giving an overall difficulty rating of easy, medium and hard. While this would certainly help students that are not as confident to take on challenges that are more appropriate for themselves, this does nothing to tell students about which level of difficulty would be the most appropriate choice to challenge themselves enough to learn as they progress. This categorical system only gives a difficulty comparison to a generalized normal, and not in comparison with the participating students themselves.

One potential method to address this problem would be to adapt a rating system to the students and the challenges. In this sense, this would be a system where each student is given a rating that is a direct indication of their skill level. A simple method that could potentially be implemented would be one that borrows inspiration from an existing rating system such as the Elo system. The Elo system is designed to rate the skill of players in a one-on-one competitive environment, specifically taking into consideration the likelihood that each player has to win [1]. While designed for use in the ratings of competitive chess players, the Elo system is also commonly adopted among many modern applications such as ranking professional basketball players or players in online competitive multiplayer games. Using this system in an educational environment would allow for appropriately judging the skill of a student, then giving the student the opportunity to engage in equally rated work that will benefit them the most.

4. Giving a Rating

In adapting a rating system to a challenge based program, it is necessary to give both the participating students and the challenges themselves a rating. Students and challenges are then considered as opponents that are competing against each other.

While both parties need to receive a rating for the same use, each must be rated in a different manner.

4.1. Challenge Rating

The challenges themselves are somewhat simple to rate, as cyber security challenges are typically designed with a specific vulnerability in mind. As such these challenges are created with a relative difficulty that can be initially applied as a rating. As a general example, a challenge that requires no specific tools should naturally be rated less difficult than one that does require interaction with a tool. Of course this still only provides a generic rating of the difficulty as described before. To overcome this, the program will essentially need a set of training data to gauge the perceived difficulty of the available challenges. This will likely have to be collected from the first set of students that would interact with the program, as we do not want the influence of more skilled teachers or those who have worked on creating the project to influence the data. The challenge difficulties collected at the end of a class that interacts with the program could be used as an initial rating, which would then be adjusted for subsequent classes. The methods of adjusting this difficulty will be discussed more in section 5.

4.2. Student Rating

The students themselves provide a larger problem of being given an initial rating since their ratings could vary much more than the challenges. A typical cyber security class could vary in students that have no prior knowledge of the course material to students that have self studied some of the material, such as students involved in a cyber security club. Separating these two cases can allow the more advanced students to work on challenges that are appropriate for their skill level while allowing the beginners to ease themselves into the material. Since programs like JagWarz are designed to be implemented in the classroom, implementing an initial survey would likely give the best indication of student skill level. This could be done in a number of different manners such as giving students a small collection of challenges and seeing what the students can complete, or giving students a paper test that their score will then correlate to their initial rating. Unlike the challenges, the initial rating of the students is not as much of a concern. In a case such as a student getting rated higher than their actual skill, the student still has the ability to work on challenges that are at a slightly lower skill level than their own rating to increase their knowledge.

Similarly students that receive a lower rating than they feel appropriate will simply be able to progress quickly to a more appropriate rating.

5. Rating Models

Since the proposed idea of facing students against challenges is not modeled in the same way that rating systems match opponents, we have designed two different models for adjusting the ratings of both the challenges and the students. The first involves adjusting the ratings of the challenges on a per student basis, while the second discusses making each challenge an opponent of the entire class.

5.1. Individual Vs. Challenge

The goal of the IVC model is for the ratings of the student and the challenges to be indicative of each students individual performance. To do this it is necessary that rather than the challenges have an overall difficulty, they will have an instance of their difficulty rating for each student involved in the class. With the IVC model, each student-challenge interaction can only have one outcome: the student either completes the challenge or is unable to do so. The adjusting of both the challenges and students rating in this case is straightforward. In the case that the student successfully completes the challenge, the students rating is increased and the challenges difficulty is decreased in relation to the student. In the other case that the student is unsuccessful, we simply need to raise the rating of the challenge. A visual depiction of this model can be seen in figure 1.

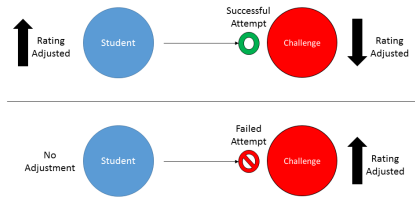


Figure 1. Individual Vs. Challenge Model

We have also developed a generic method that can be used in adjusting both the rating of the students and the challenges in this model. In the following functions, R_s stands for a students rating, R_c for a challenges rating, and $P(S)$ for the probability index of a student successfully completing a

challenge. The probability index is derived from the difference in the students rating and the challenge rating. Students that have a small difference in rating from the challenge will be low on the index, and students that have a larger gap are placed higher on the index. For the instance in which a student is able to complete a challenge, the ratings are adjusted as follows:

$$R'_s = R_c + P(S)$$

$$R'_c = R_s + P(S)$$

In the event that a student is unlikely to complete a challenge, they will be rewarded for being successful by receiving a boost to their rating due to their higher placement on the probability index. The adjustment of the challenges rating also takes into consideration that the said student was unlikely to succeed by lowering amount that the rating will change by. Essentially this means that even if a student completes a challenge that they were not expected to, it is likely that they would still find this challenge difficult were they to attempt it again. Thus the adjusted challenge difficulty should still be indicative of the students skill level. One consideration to be made is that students that have a large enough gap in their rating from the challenge should not be allowed to attempt the challenge. This will prevent students from making large leaps in rating and skipping over some material entirely. The gap that is considered too large can be determined as a rating gap that does not fall within the probability index.

In the case that a student is unsuccessful at completing a challenge, the rating of the challenge is adjusted by the following, where $P(F)$ is the probability index of a student failing to complete the challenge:

$$R'_c = R_c + P(F)$$

In this case, the failure probability index is essentially the reverse of the success probability index. For this index students that have a large difference in rating from the challenge will be placed lower on the index and students with a small gap placed higher. The main goal of this reversed index is to accommodate for unexpected failures. If a student failed a challenge that they were expected to pass based on their gap in rating, this would indicate that the challenge was actually more difficult than the rating indicated. It would then be necessary to adjust the challenges rating upwards. Conversely if a student fails a challenge that is a decent margin above their own skill level, then it is unnecessary to adjust the challenge rating upwards as much since it was already expected for the student to fail.

As an example, consider a student with a rating of $R_s = 20$, a challenge with rating $R_c = 25$, and a probability index as follows:

| | | | |
|---------------------|------|--------|-----|
| Rating Gap | 0-2 | 3-5 | 6-8 |
| Success Probability | High | Medium | Low |
| Rating Bonus | 1 | 2 | 3 |

We are then able to adjust the student rating and the challenge rating according to the method as such:

$$R'_s = 25 + 2 = 27$$

$$R'_c = 20 + 2 = 22$$

In this example the student is able to move up by an amount that should feel representative of how difficult it was for them to complete the challenge. The amount of change that feels appropriate can be determined by the range of possible ratings and the number of available challenges. Optimally, students will be able to move up in increments that place them within a margin of the next highest challenge that should be difficult enough for them. The rating range and probability index can both be adjusted so that ratings are not altered too quickly or by too much at once.

5.2. Class Vs. Challenge

The CVC model is oriented at determining the overall skill of a group of students rather than the individuals. In the CVC model, each challenge will monitor the students attempts to complete said challenges. A visual representation of this model can be seen in figure 2.

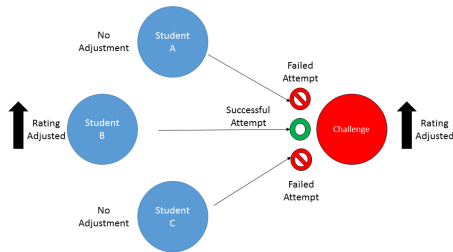


Figure 2. Class Vs. Challenge Model

Each student attempt can only have four specific outcomes based on whether or not the student is expected to be successful or not. When the student is likely to complete the challenge there can either be an expected successful attempt or an unexpected failed attempt. When the student is unlikely to complete the challenge the options become an expected

failed attempt or an unexpected successful attempt. It is then possible to adjust the difficulty rating of the challenge based on the number of students that fall into each of the four categories over a short period of time. This would be accomplished by giving a weighting to each of the possible outcomes. For outcomes where the outcome is as expected, the weight would be small and have a small effect on the rating of the challenge. In outcomes where the outcome is unexpected, we can give a higher weight to these outcomes which will in turn have a greater effect on the rating adjustment. Outcomes where students are successful are giving a negative weighting, and unsuccessful outcomes are giving a positive weighting. This then allows for the averaged outcome to show whether the challenge rating should be increased or lowered. The weighting can be then modeled by the following, where n represents one of the four outcomes, Θ represents the weight of the current outcome, and S represents the number of students that resulted in the current outcome.

$$\Delta = \sum_{n=1}^4 \Theta_n(S_n)$$

$$R'_c = R_c + \Delta$$

The rating of the challenge is then able to be entirely determined by the overall success of the students, and cannot be influenced by outlier students that are at a much higher or lower rating. The rating of the students themselves can simply be updated by use of the IVC model.

As an example, consider a challenge that has collected the following results:

| Outcome | Θ | S |
|--------------------|----------|-----|
| Expected Success | -0.25 | 13 |
| Unexpected Success | -0.5 | 3 |
| Expected Failure | +0.25 | 7 |
| Unexpected Failure | +0.5 | 2 |

This would result in $\Delta = -2$, which is appropriate as most of the students had a result that would result in a small lowering of the difficulty. This change is slightly offset by the smaller amount of students that had outcomes which increased the rating. As a result the overall difficulty of the challenge is not changed dramatically, and is representative of the entire class. The weighting of each of the outcomes can also be changed as necessary similar to the variables of the IVC model.

5.3. Student Rating Adjustment

While adjusting the ratings of the challenges requires some special considerations, the students

ratings are more simplified. Unlike the Elo system, the students ratings should never be adjusted downwards. The Elo system accommodates for the fact that the individual skill of a player may decline over time, or that the relative skill of a player will be lowered when stronger players are introduced as new opponents. However in a classroom setting where this program would be implemented, students should only be progressing in skill rather than lowering. Since the goal is for students to be motivated to progress through the challenges, we do not want to lower students individual ratings. If we were to lower the ratings of students this might cause them to be discouraged and lose motivation towards working on the challenges [2]. Without lowering the ratings of the students, the students are placed in a sort of progression system as they can only move higher in rating over time. This is also able to more effectively show students their progress over time, as at the end of the program they will be able to see how high their rating moved in comparison to themselves and other classmates.

6. Survey

One consideration to be made toward using a rating system is that the ratings of both the students and the challenges are not absolute. In competitions where such rating systems are used, a high rating is directly related to the overall skill of that player. In the classroom setting this is not the same case, as the students ratings are indicative of their readiness to take on similarly rated challenges along with their progression through the class. As such, it is critical that the students feel that their own rating and also the rating of the challenges are appropriate. The best way to do so would be to survey the students directly to receive their feedback on the system.

As students are likely not eager to participate in a survey, we would like to collect as much information as possible from the students interactions with the system without directly asking the student. For example, a data point of interest would be the difference in skill level between a student and a challenge they select. This would give us a direct look at what students may consider to be a fair matchup, as it is unlikely that a student would take on a challenge that they would consider to be more difficult than they feel ready for. This would also allow to identify whether or not students are choosing to work on challenges that are below their skill level. This could indicate that students might feel that their rating was placed too high, and they do not feel that they are on the level that their rating indicates. Another possibility would be to compare

this difference in skill rating to the amount of time it takes a student to complete a challenge. This could possibly give more insight to a specific challenges actual difficulty rather than its perceived difficulty. Other potential data points include: number of failed attempts per challenge, rate of change in student rating, difference in rating at the beginning of the course vs. the end of the course, and average rating of all of the students within the class.

Of course it is also necessary to ask for direct feedback from the students of their thoughts on the program. The main goal of this feedback would be to determine the general attitude that the students have towards the system. We want the students to feel that they are correctly ranked, which will then allow them to be motivated to progress through the challenges [5]. This information can only be determined by the students rather than the analytical data mentioned previously. Rather than asking students directly whether or not they felt they were matched appropriately it would be more beneficial to derive this answer from tangential questions. For example if we were to ask students: "What is the highest rating of a challenge that you would feel confident to attempt?", this would give insight into how appropriate a student feels their ranking is. If their response is within a small deviation of their own ranking, we can then determine that the said student is appropriately matched. Conversely if the student gives a response of a challenge rating that is far above or below their own rating this would indicate that this student does not feel that their rating is appropriate for their own skill level. Another potential question would be: "How far do you feel that you have progressed since the start of this course?", where the response is on a scale of not very much to very much. This response can then be directly compared to the difference in the students change in rating during the entirety of the course. We would then be able to determine if students that had a strong response also had a large change in rating, which would indicate that they had actually progressed and were able to also feel that they had done so. This would also help in identifying students that had a relatively small increase in rating that feel that they still had a significant improvement, which indicates that they were highly motivated even when struggling to improve.

7. Future Work

Future work includes adapting this system to an existing challenge program. As mentioned before JagWarz would be a candidate, but this concept could be applied to any program that provides a

variety of challenges to work with. Furthermore this system could be extended to the cyber security challenges that such programs are inspired from. Adapting the rating system to a competition environment could possibly create skill leagues where participants are both working on challenges that are appropriate for their skill level while simultaneously competing against others that match their skill level. Rating systems could also be extended to team based capture the flags, where teams are placed against one another within a marginal skill level to allow closer competitions.

As for the challenges themselves, one problem is that it may be difficult to provide students with enough challenges to work with, as in a classroom setting where a program is implemented both topics covered and time available are limited. As such it could also be difficult with providing a range of difficulty that allows for a natural progression for the students if there is not enough for them to work with. To address this issue it could be possible to break up challenges into steps that each have their own rating. The students would then be able to progress through the parts of the challenge as they are ready for it, thus completing the overarching challenge once they have finished each individual part. This would also provide a larger amount of data points to collect information from, which helps with the ratings of the challenges as described.

8. Conclusion

In this work we have addressed the problem of properly matching students in a cyber security education setting with work that will properly challenge and motivate students to succeed. Utilizing the combination of challenge based education programs with a modern rating system will allow for students to be appropriately matched with work that is suited for their skill level. Students then will be able to progress through the course material in a natural manner as they are able to complete challenges that offer them the correct amount of difficulty. With the two proposed models students will be able to track their own progress, and teachers will be easily able to monitor the progress of the class as a whole.

Acknowledgments

This research was made possible with the support of the Indiana University-Purdue University Indianapolis Department of Computer Information and Information Science, as well as through funding from the National Science Foundation and the United States Department of Defense. The author

would like to thank Dr. Feng Li, as well as Dr. Eugenia Fernandez, and Sheila Walter for their support

References

- [1] O. Rotou, X. Qian, and M. von Davier. (2015). *Ranking Systems Used in Gaming Assessments and/or Competitive Games*. Report. Available: <http://www.ets.org/Media/Research/pdf/RM-15-03.pdf> [July 22, 2016].
- [2] S. Fulton. "What are we teaching in cyber competitions?" in *2012 Frontiers in Education Conference Proceedings*, October, 2012. pp. 1-5.
- [3] K. Stewart, J. Humphries, and T. Anel. "Developing a Virtualization Platform for Courses in Networking, Systems Administration and Cyber Security Education." in *Proceedings of the 2009 Spring Simulation Multiconference*, March, 2009.
- [4] H. Kavak et al. (April 2016). *A Characterization of Cybersecurity Simulation Scenarios*. Conference Paper. Available: https://www.researchgate.net/publication/299820368_A_Characterization_of_Cybersecurity_Simulation_Scenarios [July, 15, 2016].
- [5] T. Flushman et al. "This is not a game: early observations on using alternate reality games for teaching security concepts to first-year undergraduates." in *CSET'15 Proceedings of the 8th USENIX Conference on Cyber Security Experimentation and Test*, August, 2015. pp. 1-1.