Adversarial Teaching Approach to Cybersecurity: A Mathematical Model Explains Why It Works Well

Christian Servin[1], Olga Kosheleva[2], and
Vladik Kreinovich[3]

[1]El Paso Community College, 919 Hunter Dr.
El Paso, TX 79915-1908, USA, cservin1@epcc.edu
Departments of [2]Teacher Education and [3]Computer Science
University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

1. Cybersecurity Is Important

- In the modern world, everything relies on computers.
- This is even more so with the current COVID'19 pandemic.
- Computers run our communications, control our utilities, largely control our planes, cars, etc.
- For our civilization to function, it is important to protect all these computer systems from malicious attacks.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

2. Teaching Cybersecurity Is Important

- Whatever automatic tools we place in to prevent cyber-attacks, smart adversaries learn to overcome.
- The only way to maintain cybersecurity is:
  - to train a large corpus of specialists
  - who would protect us from all the newly appearing threats.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

3. Traditional Way of Teaching

- The usual way of teaching any material is to present, to the students, the needed information and skills.
- With respect to cybersecurity, this means explaining, to the students:
  - the main types of cyber-attacks and
  - the main ways to defend against these attacks.
- After that, we can let the students show their creativity, but usually, teaching the basics is a must.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

4. Adversarial Teaching: A Successful Alternative Approach

- Interestingly, lately, a different approach has been very popular and very successful, in which:
  – instead of teaching students the usual way,
  – the instructor divides the class into one or more pairs of sparring mini-teams.
- In each pair, the teams interchangingly attack each other and defend their team from a partner's attacks.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

5. This Works, But Why?

- The above strategy works, which is somewhat surprising.
- We do not have a thorough coverage of all possible topics.
- So, one would expect gaps in the ability of students who have been taught this way.
- However, there are usually no such gaps.
- So, the first question is: why this approach works?
- A natural second question:
  - is this approach close to optimal
  - or we can drastically further improve it – and if yes, how?

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

6. What We Do in This Talk

- In this talk, we answer both questions.
- We explain why the adversarial teaching approach works.
- We also show that this approach is – in some reasonable sense – optimal.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

7. A Similar Approach Works in Design

- For teaching, this approach may be somewhat new.
- However, a similar approach works in military engineering.
- For example, new fighter planes are designed as follows.
- This design uses using a program that simulates dogfights between different planes.
- The first stage is natural:
  - we consider several possible designs, and
  - for each of them, we simulate how this design will perform against the existing planes.
- We continue doing this until we find a design that can beat all the possible opponents.
- At first glance, this may seem to be sufficient.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

8. A Similar Approach in Design (cont-d)

- However, on second thought, it is not:
  - it is not enough for a future plane to be better that what the opponent has now,
  - we need to have a design that will be better than what the opponent will have in the future.
- To design such a plane, we perform the second stage of the design process.
- Namely, we design a plane that:
  - is not only better than the current planes, but
  - also better than our first-stage design.
- Then, we design a plane that will be better than the second-stage design, etc.
- At the end, we get an almost perfect future plane.
- This is what is then implemented and tested.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

9. What Can We Conclude from This Fact

- A similar idea works successfully in such completely different application areas as:
  - teaching cybersecurity and
  - designing fighter planes.
- This makes us confident that these successes are not due to any specific features of these areas.
- These successes are due to the general structure of this approach.
- Let us therefore describe a simple mathematical model that would capture this structure.
- We are not specialists in plane design.
- As educators, we are clearly more familiar with educational applications,
- So, we will illustrate it on the example of teaching.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

10. Towards a Model

- We want the students to be able to handle all possible attack situations.
- Of course, different situations are all somewhat different.
- Ideally, what we want is to make sure that:
  - whatever new situation surfaces,
  - the students should have some experience successfully fighting a similar attack in the past,
  - this experience would help the student fight the new attack as well.
- In mathematics, a natural way to describe similarity is by a metric $d(a, b)$ on the set S of possible situations.
- This metric describes to what extent situations a and b are different from each other – or similar to each other.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

11. Towards a Model (cont-d)

- The smaller the distance $d(a, b)$, the more similar are situations a and b.
- In these terms, "similar" means that the distance $d(a, b)$ is $\le$ some small threshold value $\varepsilon > 0$.
- Therefore, we arrive at the following model.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

12. Resulting Model

- We have a set S of possible situations.
- On this set, we have a metric $d(a, b)$.
- We want the student to experience situations $s_1, \ldots, s_n$ such that every situation s from the set S is ε-close to
- In mathematics, such a set is known as an ε-*net*.
- The exact value of the threshold is determined by our resources.
- The smaller ε, the better.
- However, a drastic decrease in ε would mean a drastic increase in situations experienced during teaching.
- And the teaching time is limited.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

13. How Do We Compare Quality of Different Teaching Schemes

- Once we fix $\varepsilon > 0$, a natural measure of quality is the number of experiences situations n.
- The smaller n, the faster we can train.
- Alternatively, we can fix n – and thus, the training time.
- Then, we need to find the situations $s_1, \ldots, s_n$ that lead to the smallest possible ε.
- For each metric space, the smallest possible number of elements in an ε-net is called ε-*entropy*.
- To be more precise, usually the logarithm of this smallest number is called the ε-entropy.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

14. The Corresponding Optimization Problem Is NP-Hard

- It is known that problem of finding the smallest ε-net is, in general, NP-hard.
- This means, crudely speaking, that:
  - unless P = NP (which most computer scientists believe to be false),
  - no feasible algorithm is possible that would always find the optimal ε-net.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

15. Let Us Reformulate Adversarial Teaching in These Terms

- The first team starts with some attack situation s_1.
- Then, the sparring team learns how to defend against this attack.
- So, next time, the attacking team will try to find:
  - a new way of attacking that has the most chances of success,
  - i.e., the situation s_2 which is as far away from the original situation s_1 as possible:

$$d(s_2, s_1) = \max_{s \in S} d(s, s_1).$$

- Then, the sparring team learns how to deal with the situation s_2 as well.
- The next attacking situation s_3 will be as far away from both s_1 and s_2 as possible.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

16. Adversarial Teaching (cont-d)

- So, the distance $d(s, \{s_1, s_2\}) \stackrel{\text{def}}{=} \min(d(s, s_1), d(s, s_2))$ is the smallest possible:

$$\min(d(s_3, s_1), d(s_3, s_2)) = \max_{s \in S} (\min(d(s, s_1), d(s, s_2))).$$

- In general, once we have experienced the situations $s_1, \ldots, s_k$, we select the next situation s_{k+1} for which

$$\min(d(s_k, s_1), \ldots, d(s_k, s_{k-1})) =$$

$$\max_{s \in S} (\min(d(s, s_1), \ldots, d(s, s_{k-1}))).$$

- We continue while there is a situation which is different from all the previous ones: $d(s_k, s_i) > \varepsilon$ for all $i < k$.
- When this is no longer possible, we stop; then:

$$\max_{s \in S} (\min(d(s, s_1), \ldots, d(s, s_n))) \leq \varepsilon.$$

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

17. This Strategy Works: A Proof

- There are only finitely many possible situation.
- Indeed, each situation has to be described in a reasonable time.
- Thus, it contains a reasonable number of characters N to describe.
- For each N and for each set of possible symbols, we have a finite number of strings of length $\le N$.
- At each iteration, we generate a situation which different from all the previous once.
- Thus, eventually, the above process will stop, and we'll have $\max\limits_{s\in S}\left(\min(d(s,s_1),\ldots,d(s,s_n))\right)\le\varepsilon$.
- This means that every situation $s\in S$ is ε-close to one of the situations s_i.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

18. This Strategy Is Asymptotically Optimal: Formulation

- Let n be the number of situations that the students have experienced by following this strategy.
- The strategy is feasible.
- However, the problem is NP-hard.
- So, we cannot expect that for this number n, the threshold ε is optimal.
- It is thus possible that, in principle, with the same number n, we can reach a smaller value ε'.
- What we *can* prove, however, is that this decrease cannot be too drastic; namely:
  - even for one fewer $(n-1)$ situation,
  - the corresponding optimal value ε' is at best twice smaller, i.e., that $\varepsilon' \geq \varepsilon/2$.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

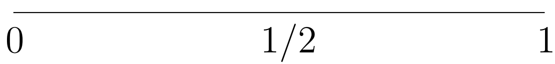
19. This Strategy Is Asymptotically Optimal: A Proof

- Let us prove this optimality result by contradiction.
- Indeed, by our construction, we have $d(s_i, s_j) \geq \varepsilon$ for all $i \neq j$.
- Suppose that we have a ε'-net $s'_1, \ldots, s'_{n-1}$.
- By definition of a ε'-net, each element s_i is ε'-close to some element $s'_{e(i)}$.
- For $i \neq j$, we cannot have $e(i) = e(j)$: otherwise, we will have $d(s_i, s_j) \leq d(s_i, s_{e_i}) + d(s_j, s_{e_i}) \leq 2\varepsilon' < \varepsilon$.
- Thus, to each of the n elements s_i, we assign a different element s'_j.
- However, this is impossible, since we assumed that we only have $n - 1$ elements e'_j.
- The optimality is thus proven.

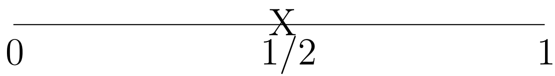
Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

20. Graphical Illustration

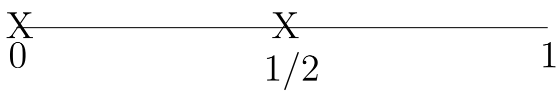
- To make it easier to understand, let us give two simple geometric illustrations of the above idea.
- Let us start with the simplest example of a metric space S – namely, the interval $[0, 1]$:



- It is reasonable to select the midpoint $1/2$ as s_1:



- There are two points that are the farthest from s_1: the left endpoint 0 and the right endpoint 1.
- Without losing generality, let us select $s_2 = 0$:

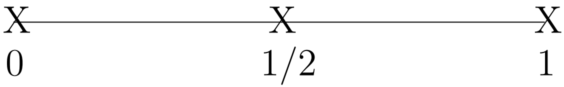


Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

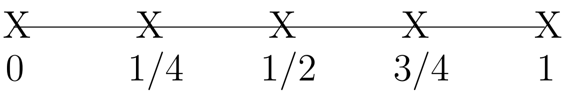
21. Graphical Illustration (cont-d)

- Now, $s_3 = 1$ is the point with the largest value of

$$d(s, \{s_1, s_2\}) = \min(d(s, s_1), d(s, s_2)) :$$



- At this stage, the midpoints between 0 and 1/2 and between 1/2 and 1 are the farthest from the set $\{s_1, s_2, s_3\} = \{0, 1/2, 1\}$.
- So, after two stages, we add them both:



- Now, the largest possible value of $d(s, \{s_1, s_2, s_3, s_4, s_5\}) = d(s, \{0, 1/4, 1/2, 3/4, 1\})$ is $1/8$.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration
Home Page
Title Page
Go Back
Full Screen
Close
Quit

22. Graphical Illustration (cont-d)

- So, at the next stage, we add one of the points in between the existing ones, e.g., the first one (1/8):

X X X X X X

0 1/8 1/4 1/2 3/4 1

- After three more stages, we add all midpoints, so we arrive at the following configuration:

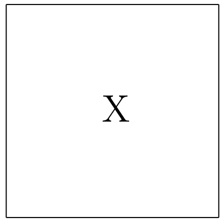
X X X X X X X X X

0 1/8,1/4,3/8,1/2,5/8,3/4,7/8, 1

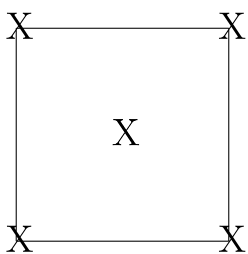
Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration
Home Page
Title Page
Page 23 of 29
Go Back
Full Screen
Close
Quit

23. 2D Example: Square

- For a unit square, we get a similar situation.
- First, let us pick the midpoint as s_1:

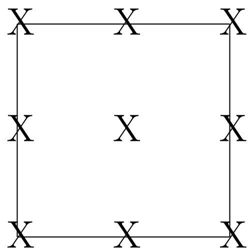


- Then, the next four selections s_i are the vertices:

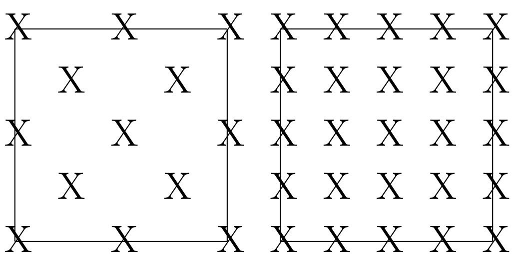


24. 2D Example: Square (cont-d)

- After this, the next four selected points s_i are he mid-points of the four edges:



- Here, we have, in effect, four sub-squares.
- On the next stage, the same procedure is repeated for each sub-square, etc.



Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

Home Page
Title Page
Page 25 of 29
Go Back
Full Screen
Close
Quit

25. What We Did

- We provided a *simplified* mathematical model that explains why adversarial teaching works.
- We showed that, in some reasonable sense, adversarial teaching is indeed a close-to-optimal teaching strategy.
- The existence of such an explanation made us more confident that this method is a right one.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

26. Can We Do Better?

- Teaching with more confidence is good.
- However, it would nice to have a model that helps us teach *better*.
- For this, we need a more realistic model.
- Such model should take into account that:
  - some attacks are more difficult to defend against, while
  - other attacks are easier are easier to defend.
- Such models should take into account team dynamics.
- We hope that our simplified model will provide a starting point for developing such more realistic models.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

27. How to Motivate?

- In this talk, we concentrated on the technical part, on *what* to teach.
- We implicitly assumed that students have the needed motivation (and, of course, the needed background).
- In reality:
  - while some students are always eager to learn,
  - for other students, it is important to keep them motivated.
- In our experience, when properly organized, competitive environments like hackathons are great motivators.
- But pedagogy teaches us that many students do not perform well in competitive environments.
- How best to motivate is still an open problem.

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration

28. Acknowledgments

This work was supported in part by the National Science Foundation grants:

- 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science),
- HRD-1242122 (Cyber-ShARE Center of Excellence).

Teaching . . .
Adversarial Teaching: . . .
This Works, But Why?
Towards a Model
Resulting Model
The Corresponding . . .
This Strategy Works: . . .
This Strategy Is . . .
Graphical Illustration