8-2020

Analyzing Network Topology for DDoS Mitigation Using the Abelian Sandpile Model

Bhavana Panchumarthi
Reed College

Monroe Ame Stephenson
Reed College

Monroe Stephenson

Hey this is monroe and with me is bhavana

Bhavana Panchumarthi

Hi

Monroe Stephenson

We are both participants of the altREU internship at psu welcome to our podcast: Sandpiles the master of DDoS. In this podcast, we hope to share our research on combating DDos Using the mathematical modeling tool called the abelian sandpile model.

Bhavana Panchumarthi

For this project we had Dr. Art Duval of UTEP work with us, who's here for an interview. We are also going to be interviewing Saketh whos starting as a freshman at UC Berkeley

Monroe Stephenson

If i'm attacked by some cyber attack, whatever it may be how am I supposed to know when i'm attacked?

Bhavana Panchumarthi

Right, if you're especially someone who uses their web browser to do basic things like browsing netflix or emailing or doing some online shopping, a cyber attack may seem abstract without a tangible attacker.

Monroe Stephenson

Let’s take a DDoS attack for example. Which ddos stands for distributed denial of service. When something is ddosed there is an unavailable error because your computer cannot access the server to get the information

Bhavana Panchumarthi

Yeah and that's exactly what you see when you try to access a website but you end up looking at service unavailable error and so this error can be due to unexpected overloading, but when it is due to a ddos attack it is much more serious and the server that you are trying to reach out is experiencing this attack

Monroe Stephenson

And the disruptions that internet users face due to a DDoS attack is what we are trying to minimize.

Bhavana Panchumarthi

We would like to welcome Saketh who's a rising freshman at the University of California Berkeley. He is co-founder of project morph an organization, dedicated to taking a stance against domestic violence.

Monroe Stephenson

He also has interned as a software engineer at the educational management tech company Kudos wall.

Saketh Panchumarthy

I know a little bit about DDOS. But what exactly are you guys looking into about it?

Bhavana Panchumarthi

We're looking into a method to mitigate DDOS attacks called blackholing. We're looking into blackholing because it is a common method that is easy to implement and cheap so many nonprofit and governmental organizations can use our resources to optimize their use of blackholing as a method of mitigation.

Saketh Panchumarthy

What is blackholing?

Monroe Stephenson

It is when one server is sacrificed in order to allow all the other servers to stabilize so that the network can go back online all the data that enters the blackhole gets deleted and the blackhole cannot tell the difference between the good data and the bad data or malicious.

Saketh Panchumarthy

Why would you want to do this? It seems like a bad idea because you're just throwing information away.

Bhavana Panchumarthi

Well, sometimes it's the only way to approach the issue. It doesn't take too fancy of tactics. It's relatively easy to implement and it works well enough at the end of the day. We want our Network to go back online and also save the organization, so the loss of clients

Saketh Panchumarthy

So what server is usually the best to be a blackhole?

Monroe Stephenson

Well, we've been totally looking at backbone Network. So usually the one that is attacked is the best one to be by Cold. However, that's not always the case because it can be a bit more complicated.

Saketh Panchumarthy

So what is a backbone Network and why is it so important?

Bhavana Panchumarthi

So there used to be something called the internet backbone and that's not really the case anymore because we don't have one large Network that makes up the core of the internet. We have lots of subnetworks and we call these backbone networks because they make up the internet. We as we know it today and these are usually the internet service providers. We know like AT&T or Cogent Communications or networks like that. And if they are attacked there's a lot at risk because we have a lot of people using their services.

Monroe Stephenson

So if all the backbone networks were attacked then all of the sudden networks would crash but this would take a massive attack to accomplish. many precautions have been put in place so that this doesn't happen. However, if it did it would be catastrophic to the internet. one more thing backbones networks tend to be much simpler since they're kind of a bird's eye view. So it's generally much easier to make a model. So it is generally much easier to make models of the backbone. at work

[redacted]

Our problem that we want to tackle is minimizing the amount of data that ends up in the blackhole. We want to approach this issue because less data that gets deleted within the network is better for the network in total and for the customers. The problem we are looking at is how to optimize A which server you choose to be the blackhole and B how to minimize the amount that goes into said server. We do this because we want the most amount of data and therefore customers to be content.

Bhavana Panchumarthi

To approach this optimization problem we are going to be using a mathematical model which is the Abelian Sandpile model. And we have our faculty mentor Dr. Art Duval here to help us discuss that. Welcome Art

Monroe Stephenson

Tell us a bit about yourself Art.

Art Duval

So my name is Art Duval. I'm a professor of mathematical Sciences at the University of Texas at El Paso, and I've been here since 1991. My research is in algebraic and topological combinatorics, which means that I count things. That's the combinatorics part and the topological part is that the things that I count are somehow topological or geometric in nature. One of the things that I've spent a lot of time looking at in the past several years is about what I'll call spanning trees of higher dimensional complexes. So if you take the networks, like in the problem that you're looking at a spanning tree is a collection of edges or connections that connect all the cities, but do it with the fewest edges possible. And so one of the things that we've been doing is looking at how to make that work in higher dimensions. In addition to my mathematical research. I've been involved in educational issues for about 20 years, especially around issues of teacher preparation and the surprisingly deep mathematics that underlies the mathematics that

goes on in the K through 12 schools. About a year ago I stepped down from serving as a contributing Editor to the American Mathematical Society blog on teaching and learning mathematics, which I had done for five years.

Bhavana Panchumarthi

So using his experience as our faculty member Art has guided us through using a mathematical model for our blackholing problem. So Art what advice do you have about using mathematical modeling to analyze real world scenarios?

Art Duval

Sure, you know one thing that's really great that I enjoy about working with the two of you is that you already had an idea. You knew what you wanted to solve and I think that always makes it better if people come in with an idea of what it is that they want to do and that they're working on something that they're interested in that's certainly always worked for me. And I'd say that actually I haven't had to say too much to you because you both already have so many good ideas, but I think one thing that I was able to contribute is The idea that you would see some patterns in a matrix or something like that and I was able to say well, you know, how does that row as a matrix correspond to a particular vertex of the graph? So, you know the idea about looking at your data or whatever and then go back to the mathematical model. Earlier in my career when I wanted to work on applications. I'd look for my math in their problems. I'd go talk to some about the problems. Like how can I turn this into an interesting common networks problem and that never worked ever but what works better is that now I when I talk to people about applications to say tell me about your problem and then I see if there's anything that I know that can be helpful and that's been the most useful thing in terms of making myself useful as a mathematician to people who have real problems.

Monroe Stephenson

The mathematical model we used is called the abelian sandpile model. How would you briefly explain the abelian sandpile model without any real mathematical notation?

Art Duval

Okay. Sure. So let me start with a particular model. Imagine. You have a big grid. It could be a square grid. It could be a hexagon grid, it could be any sort of grid and so what does that mean? That's a bunch of you know vertices like cities in your network and edges that connect the pairs of vertices. And so on this grid we're going to put grains of sand integer numbers of grains of sand on on the vertices and we call such an arrangement like that, especially it's a configuration. And what we're going to do is we start with some configuration. Then we're going to add a grain of sand one at a time to a particular spot a particular vertex. and then if any vertex has too much sand it's going to topple to its neighbors and what that means is if the number of grains of sand is at least the number of neighbors, then that vertex will send one grain of sand to each of its neighbors. And so that's all well and good. But then as the first vertex sends sand to its neighbors those neighbors might now suddenly have too much sand and they will also topple and this just keeps going until it stabilizes until no vertex has too much sand. The abelian in the abelian sandpile model is because you might wonder well if I have two different vertices that both have enough sand to topple. Which one should I do first? Abelian is a word we use in math to mean that the order of things doesn't matter and that's what's going on here the order in which you topple the vertices doesn't matter. And I said to imagine this on a big grid but actually you can do this on any graph or network just like the networks that you're looking at and so in general this is just any collection of vertices and they're all connected by edges. Any way you like they don't even you don't even have to be able to draw them. in a plane There's two interesting things. I want to say about the abelian sandpile model one is it was initially introduced as an example of something called? self-organizing criticality self-organizing part is about how each vertex just has its one little local rule. And we don't have to tune the parameters or anything. It just goes off on its own. The criticality is a little harder to explain but it's an interesting phenomenon and it's

related to the power law we sometimes see on large Networks. The other thing I wanted to mention is because it's near and dear to my heart is I mentioned about these spanning trees and the number of spanning trees is actually the number of certain configurations, but I won't go into the details about that.

Monroe Stephenson

Thank you. And so from we've been taking data from a data set called the internet topology zoo and we take the backbone Network and think that Network as an abelian sandpile model and just like our said we imagine that each server is a platform and that every time a server gets a request like ours said ass and gram is added and you just keep adding sand grains and you get a sand pile. However, at one point when you add too many sand grains the set and power will topple and send extra grains to the neighbouring sand neighboring servers and that is when they become overloaded.

Bhavana Panchumarthi

And so a network with multiple servers can be thought about as having different sandpiles of different sizes. And so the network is at a state where it cannot accept any more sand grains without any of its sandpiles toppling. Then you say that the network is running at a maximum capacity. During a DDOS attack. There is an overwhelming amount of malicious traffic. We get a non-desired influx in sand grains that threaten the stability of our Network. Because the legitimate sand grains and the malicious sand grains are indistinguishable. The excess sand grains do not have anywhere to go. So we want to find the best server to blackhole so that it welcomes any amount of access data and deletes it. So we're going to be calling that server a sink. Our sink is a term that we are borrowing from the brilliant sandpile model. Could you tell us about the role of a sink in a general sandpile graph?

Art Duval

Sure, it's pretty much like what you are saying that if we have too much sand in the model. Then it would actually never stabilize you just keep toppling and toppling and toppling and and we'd never hit stability ever. So the way around this is we pick one vertex and we call it the sink. And then that vertex behaves differently than all the other vertices any grains of sand that land in that vertex. They simply disappear. And that's that sink vertex. It's not allowed to top all because well it never never piles up. And so it's a remarkable thing that if you have used just that one sink vertex. I guess we have to assume a few basic things about the graph if we have just that one sink vertex, then the sample model will always stabilize. And on if you think about really big models for instance. I remember I started with talking about, you know, say a large grid what you can do is just put a single sink vertex and attach it to all the vertices at the edge of that grid and then that acts as if there's a giant I don't know circular trough around the edge of your graph and so when things get far enough away from the center, they just disappear. And so that that's what the sink does and I think it's really appropriate for your model.

Monroe Stephenson

What we want from our sink is to have the least number of sand grains in it as possible on top of considering how many grains within the network end up in the sink. Also consider how many customers are denied service for the duration that the network is down. Our model can also consider the length of the DDOS attack as a parameter while suggesting the right server to enable the sink or model is quite simplistic and built on quite a few assumptions. How can a toy model like if this be helpful for real-world networks

Art Duval

Oh, that's that's a great question. And we use toy models and Mathematics all the time. In fact, you can even think of the abelian sandpile model itself. It is really a very simple model. I'm sure real sand doesn't say how tall am I and now I'm going to topple my neighbors and yet the abelian sandpile model yields very rich results that we haven't even begun to touch here. Something about toy models in general is you have to start somewhere, right? You've got to start building your model from something and it's really hard to build a complicated model without testing it at some smaller version of itself first. But actually I

don't want to think about a smaller version. I want to think about a simpler version, a less detailed model. In this case. It might be even just also less complicated computer code. any complicated code that I've ever written I don't just write it from top to bottom and it works from the beginning. Now you start with let's make it work just for the easiest cases or with the lead. That's at the fewest number of parameters or whatever. And what's interesting is that in many situations even with these simpler models. You can already see interesting things. And so it's worthwhile not just to build it in tests. Okay, great. Let's move on but build it and test and see hey what's going on in this case? What interesting patterns do we see? And so first off sometimes that's just easier to deal with while you build up your just Theory, but also sometimes when you Spot these interesting things even in the simple cases you can use that to design. The next level up like you notice that this was interesting. Hey, maybe in the next iteration of the model. We should really take close attention to this particular feature that we saw in the simpler model and try to push it out that way. And then of course once you start simple, you can continue to build it up step by step but the longest journey begins with a single step and that's what the toy models are.

Bhavana Panchumarthi

Art, thank you for answering our questions and just talking about our experience of working together on this project. And yeah.

Art Duval

Sure. It was great to talk to you. I really enjoyed working with you this summer. Thanks a lot.

Bhavana Panchumarthi

Alright Saketh what's your opinion on the idea of applying this mathematical computational model to the idea of these Networks?

Saketh Panchumarthy

I've never thought of using abstract math concepts and describing these networks, but I think it's a good way to apply these ideas. I'd like to play around with this interface and see what it could do.

Monroe Stephenson

Do you think that this draws analogous to the real world?

Saketh Panchumarthy

I'm not sure if analogous is the best word to describe it but I definitely think that this is applicable because there's been a large number of DDOS attacks this year as a result of more people using internet services due to the pandemic. I think this is a great way to look at these ideas.

Bhavana Panchumarthi

With simplifications are being made that we may have overlooked.

Saketh Panchumarthy

I think maybe you could have looked at other factors and network security rather than simply Network structural features. I'd also like to see you consider things other than internet backbone Networks.

Bhavana Panchumarthi

All right Saketh. Thanks for being on this podcast with us.

Monroe Stephenson

Thank you.

Saketh Panchumarthy

No problem.

_____

Monroe Stephenson

our model we use Sage to build it because Sage is built on top of python and is accessible through Jupyter notebooks It also has the abelian sandpile model built in. and it has all the data structures that python comes with and best of all it's free.

Bhavana Panchumarthi

We used the network X to convert our graph files. We also used numpy which is for matrices. And we also use sci-fi and a whole bunch of other packages. They're all that. They can all still be used for sage. And we got our graphs that we mentioned earlier from a website called the internet topology Zoo, which is a large database or data set of like different backbone networks all over the world. Okay, so what can our model actually do? our model can take basically any network and analyze it. so when you analyze it what we can see is which servers can be most susceptible to attacks and which servers can actually be good black holes for a server that's attacked.

Monroe Stephenson

It can create a heat map from these matrices. These heat Maps. give a visualization where You can match up. the sink server with the attacked server. So what our model cannot do though? A handful of things currently, but hopefully soon you cannot tell you how to prematurely mitigate the DDOS stack by either adding servers or connections between servers.

Bhavana Panchumarthi

Okay, it also cannot compute quickly compute large networks because the algorithm has to go through each server and consider this source of attack and it also has to go through the same list of servers and consider them as a potential sink. So the time it takes for it to compute these networks to this algorithm grows at a quadratic rate. It's still quite simplistic despite all the parts we're considering.

Monroe Stephenson

So how do we hope to expand upon and improve our model? So one thing that we're trying to focus on is trying to optimize the algorithm so that the code runs quickly, hopefully. Quickly as in linear growth instead of quadratic. We plan to do this by deriving a formula for the Matrix that does not need this stabilization algorithm.

Bhavana Panchumarthi

Currently what the model can do is suggest countermeasure Okay, currently what the model can do is provide mitigation suggestions, but hopefully we'll be able to determine how to make the network safer in the future to prepare for a DDOS attack.

Monroe Stephenson

Lastly we hope to add more variables so that the model will be closer to reality. Such as the network might not be at full capacity initially or more servers can be attacked simultaneously.

Monroe Stephenson

A question that arises is who can benefit from this. So really anyone who uses the internet, but if we have to be more specific. Those who work within internet nonprofits working in security. As well as higher education institutes who have their own network. And so we're hoping that our model can address attacks that might happen on smaller scale networks such as the ones that occur on university campuses' networks Because in general they don't have as good of resources to address the issues. and we also hope that this can help the fast growing movement of Municipal broadbands where Municipal broadbands are local government run Internet service providers And again with small teams, it can be hard to identify where there are susceptible. So this tool could be used freely by them.

Bhavana Panchumarthi

how could this benefit someone who is not managing a network and is worried about their network security? This could actually benefit the user too because if you use the internet You might have run into problems before and this could help your experience with using that internet be more convenient. This can benefit the listener if we can just distribute this model to nonprofits and municipal broadbands or other local area networks that are in your area.

Monroe Stephenson

So how will this benefit people well? Like Bhavana said cyber security is important for use of the internet if we don't have security then the internet. Could go into chaos, especially on those smaller networks where Only a handful people could control the entire network.

Monroe Stephenson

all right. In the future you should be able to get access to this code. On our githubs, which could be found on my website or by emailing either one of us.

Bhavana Panchumarthi

Finally, we would like to acknowledge the people who will help us along the way.

Monroe Stephenson

Thank you Art Duval for mentoring us throughout the summer.

Bhavana Panchumarthi

Thank you. Saketh for joining us on this podcast.

Bhavana Panchumarthi

And thank you Dave Perkinson for introducing us to sandpiles.

Monroe Stephenson

And thank you to Christof and Mackenzie for making the altREU possible.

Bhavana Panchumarthi

And also just thank you to everyone at PSU and in the altREU who helped us just make it through the altREU and have all the resources we need.

Monroe Stephenson

We thank you the listeners for sitting and listening to what we had to say about our project.

Bhavana Panchumarthi

What if they were walking?