Data Mining: The Samurai of Networks

Data mining techniques are used to help detect hackers and protect data over networks.

Introduction

Many of us have a slight fear in the back of our minds as we navigate websites and send data across the internet on a daily basis, "What if I get a virus?" The consequences of receiving a virus, trojan, or some other attack in general can seem quite daunting. Perhaps your device could freeze or crash, your data could be stolen, or maybe someone renders control of your device from you, and now they are able to do whatever they want from a remote location. In any case, discovering techniques to protect against these attacks is of interest to all of us that utilize any type of network, such as the internet!

Take for example recent attacks on certain banks. Attackers attempted to steal millions of dollars from banks in West Africa and South Asia. Luckily, the banks were able to spot that the transactions were fraudulent, so no money was lost. And these were not out of the ordinary attacks. These cyber attacks happen all of the time. But these attacks highlight the fact that we need ways to detect invasions so we can protect our data!

The main component administrators are plagued with as they attempt to keep networks safe from intrusions is size. Billions of people on the planet have access to technology that connects them to networks. Imagine you were a teacher looking over a billion students taking an exam. It would be almost impossible to catch students that were cheating simply due to the sheer amount of them. Network administrators deal with this problem on a daily basis as they attempt to catch hackers, and they handle this problem through the use of Intrusion Detection Systems.

Intrusion Detection Systems are the snitches of networks. They essentially scan through everyone connected to the network, as well as all of their communications (for example sending a facebook message), and look for anything that seems out of place. If the detection system finds something wrong, they tattle right away and send out an alert so someone can fix the problem.

Researchers have been looking into Intrusion Detection Systems for decades, and they have found that Intrusion Detection Systems using data mining are the best at snitching. Data mining is the studying of large data sets to find patterns and predict outcomes. Intrusion Detection Systems use data mining to find patterns of certain communications over a network. This allows the detection system to know when something seems fishy.

This type of pattern matching is something we do all of the time as humans! Say you live with your spouse and child, and they love the chocolate-chip cookies you make. Normally they eat all of the cookies except for one, which they graciously save for you. However, one time, there are no cookies left. You automatically know this is not right, and someone is definitely in trouble for eating your cookie.

A natural question to ask, since we want the best protection for our data, is which Intrusion Detection System is the best? To answer this we need to look at the main two types of detection systems: Anomaly-Based and Misuse-Based.

Mind-Reading Detection Systems

Researchers give a brief description of an anomaly-based Intrusion Detection System, "An anomaly-based system uses the normal profile of a system or user to determine its decision making process." Anomaly-based systems study "normal" behavior of a typical user on a network. This provides them the ability to catch unexpected behaviors (anomalies) which could be threats.

These systems have the ability to detect new types of attacks. It is somewhat comparable to the "gut feeling" that humans get when something doesn't feel right.

Many anomaly-based systems are able to achieve this “gut feeling” with the use of machine learning.

Machine learning is a type of data analysis which allows systems to learn on their own, without the need for explicitly programmed directions. Machine learning is a technique which utilizes data mining. Picture someone that could read your mind, and then would do everything you wanted from them, but without having to tell them at all. Magical, right? That’s the beauty of machine learning.

Match. That. Attack

Researchers also give a brief description of misuse-based Intrusion Detection Systems, “They function by using patterns of recognized attacks or known critical points in a system to find and match known intrusions.” These systems differ from anomaly-based approaches in that they don’t “think” the way that those systems do. Tragic, I know, but these systems do still have their uses!

Misuse-based systems are also sometimes known as signature-based. Just as someone can identify if you signed something by your signature, misuse-based systems can identify potential threats based on the signature of previous ones. Think of them as the detectors with a photographic memory. If they’ve seen it, they can spot it.

There are definitely advantages to these systems, such as their insane level of accuracy. This is key. Where anomaly-based approaches fall short, misuse-based approaches step up to the plate. Misuse-based Intrusion Detection Systems produce far fewer false alarms, since they identify potential threats based off of previously known and verified threats. This accuracy is incredibly valuable. You wouldn’t want someone screaming “Fire!” in a theater if there was no fire, would you?

Unfortunately, just as misuse-based systems are great at what anomaly-based systems are not, they fall short where anomaly-based systems excel. Since these systems detect intrusions using data from previous ones, they are not able to detect new types of attacks like anomaly-based systems can. We are beginning to see the

struggles researchers have been facing when attempting to create the best Intrusion Detection System.

Mix and Match

Researchers are not blind to the shortcomings of both of these methods. In fact, researchers have discussed that between anomaly-based and misuse-based systems, neither is statistically better than the other. This has led to the creation of new types of Intrusion Detection Systems. They are known as Hybrid Systems.

The face of current research in the field, there are multiple types of Hybrid Intrusion Detection Systems. However, they are all constructed with the idea of combining abilities from misuse-based and anomaly-based approaches to get the best of both worlds. And this makes perfect sense! Potatoes and pizza are separately delicious items. But have you ever put potatoes on pizza? It's life-changing. Trust me.

One route that researchers have taken is stacking the misuse and anomaly-based systems on top of each other. In other words, they create a system that processes user behavior on a network through a misuse-based detection system, as well as an anomaly-based detection system. These have been proven to be more effective at snitching on the attackers, but they also have their drawbacks. One of those being that they produce too many false alarms. And no one really likes the boy who cried wolf.

A way researchers have been combatting this is by creating systems that extend the capabilities of a misuse-based system or anomaly-based system, rather than just combining them. For example, "learning" capabilities can be added to a misuse-based Intrusion Detection System in order to keep the low false alarm rate, but give it the ability to detect new types of attacks.

Another method is to intertwine the capabilities of each type of Intrusion Detection System. This allows the Hybrid System to get the abilities it needs from both, and since the systems are not just stacked on each other, the shortcomings of each don't

present themselves. In other words, there will be a low false alarm rate, and the system will be able to recognize new types of attacks.

Researchers are convinced that these are the superior types of Intrusion Detection Systems, but like everything in life, they have their flaws. One of those being the time it takes them to search a network. It's important to have great accuracy in detecting threats, but it's also important to discover them in time to do something about it. What good is the police if they can't get there in time?

Conclusion

Overall, it is evident that data mining techniques allow Intrusion Detection Systems to effectively catch malicious behavior over networks. It appears that Hybrid Systems are the future of research in this area, as researchers have found that they are superior to misuse-based and anomaly-based detection systems. Tactics need to be discovered which can make these systems run quickly enough for them to be truly beneficial. If this is done, the rewards can be beneficial to us all in many aspects of our life. For example, a Hybrid Intrusion Detection System has been presented for power systems! Their work could help prevent attacks on power grids all over the globe! Once we remember how prevalent computers and networks are in our world, the significance of this research becomes clear.