1)What is the problem?

2) What is the motivation? Why is it important?

3) What makes this approach novel?

4)What are the challenges? Why is it hard?

5) What is the solution?

6)How was the approach validated?

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| article | The problem | The motivation | | Why novel? | | What are the Challenges | | | What is the solution | | | | How was the approach validated? | Unfamiliar concepts |
|  |  |  | |  | |  | | |  | | | |  |  |
|  |  |  | |  | |  | | |  | | | |  |  |
|  |  |  | |  | |  | | |  | | | |  |  |
|  |  |  | |  | |  | | |  | | | |  |  |
|  |  |  | |  | | |  | | |  | |  | |  |
| Machine learning techniques for intrusion detection | Most techniques used in today’s IDS (intrusion detection system) are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. | Better adaptive methods like various techniques of machine learning can result in higher detection rates, lower false alarm rates and reasonable computation and communication costs. | |  | |  | | |  | | | |  |  |
| Last year | Network is open to all kinds of activities.  Thus, to verify every single communication taking place is an impossible task.  This nature of network determines that it is vulnerable to attacks.  The # and types of attacks increase.  Better way of identifying attack patterns is necessary. | | The university encompasses all kinds of individuals. So does its network hosts all sorts of activities. Moreover, data stored on the university network is both sensitive and significant.  Therefore, it is very important for the university to design a secure system that rejects cyber intrusion. | |  | | | No ground truth to test the results. Therefore, would be nice to have labeled data, that is, data with identified groups and flags for attacks.  Better hardware and algos may allow real-time/live testing of the methods mentioned in this paper.  48min of network data is adequate but still not large enough. Larger dataset may give better accuracy.  The varying/unstable/fleeting nature of attacks and malicious communication also lead to a more complicated validation process. Moreover, the way of attacking as well as the number of attacks are always changing.  Therefore, to distinguish such a volatile state from normal activity is challenging. | | | Assume that the access pattern of an attacker fundamentally differs from that of benign users.  Thus, may be able to differentiate the normal activity of an attacker vs benign user.  Specifically, logic flows:   1. All network communication 🡪 a single descriptive data point; 2. Features are extracted from this single descriptive data point. 3. Such features are then clustered to determine the different access patterns and separate types of communications. 4. These features extracted from netflow 🡪 features that summarize all the network activity of a single IP node. 5. All IP nodes are then clustered using IP level information, and such clustering should enable us to differentiate between attackers and benign users. | | Data source from network traffic of UVA, so the source of data is validated.  The approach, however, is not veritable because there is no ground truth to test the results upon. | Attack vectors  Unsupervised clustering technique  Host-source pair  How does time component work in this approach?  Hierarchical clustering algorithm.  “computers on network that switch context over time”, what is “switch context”? |