Techniques for Anomaly Detection in Network Flows

Bianca I. Colón-Rosado, Humberto Ortiz-Zuazaga

University of Puerto Rico - Río Piedras Campus
Computer Science Department

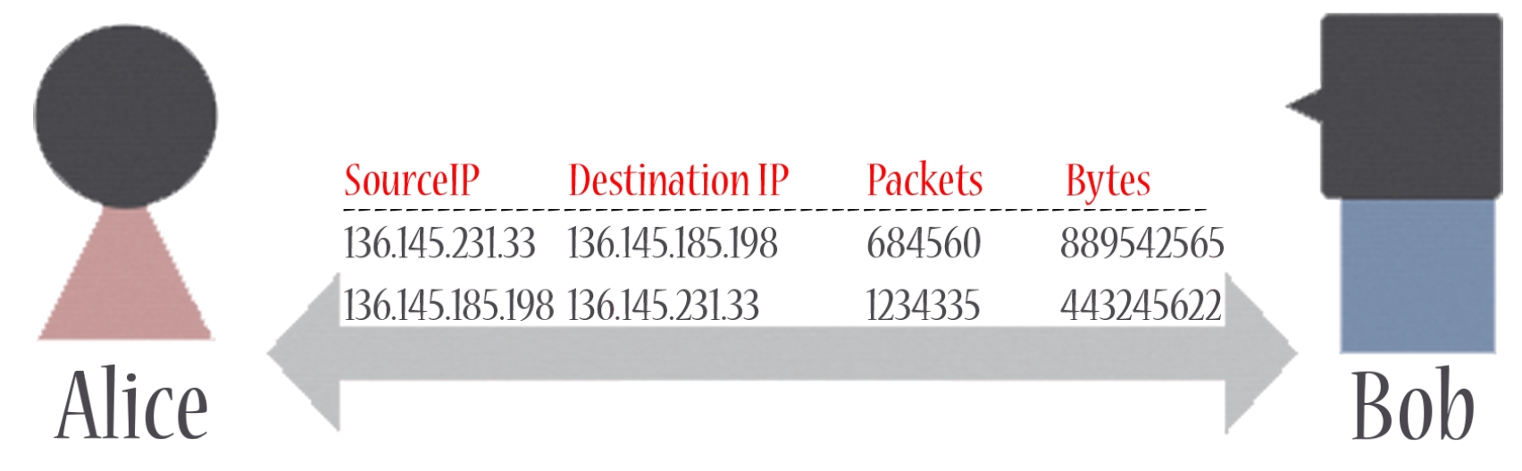
Abstract

A general method for detecting anomalies in network traffic is an important unresolved problem. Using Network Flows it should be possible to observe most anomaly types by inspecting traffic flows. However, to date, researchers are still struggling to find an effective and lightweight method. We collected one week of flow data using SiLK from the UPR's Science DMZ, a high-performance network for data science. We analized the flows with FlowBAT. No real anomaly was detected, just a false positive. We need to collect more flow data to establish patterns and find anomalies.

Introduction

This work seeks to detect anomalies in network traffic by inspecting the network flow data. For our purposes, a **computer network** is a telecommunication network that allows computers to exchange data, which is transferred in the form of packets. An **anomaly** is something that deviates from what is standard, normal, or expected.

A **network flow** is a summary of a sequence of packets sent from a source computer to a destination, which may be another host, a multicast group, or a broadcast domain. A flow could consist of source IP, destination IP, source port, destination port, packets, bytes, flags, source time, etc.



SourceIP	Destination IP	Packets	Bytes
136.145.231.33	136.145.185.198	684560	889542565
136.145.185.198	136.145.231.33	1234335	443245622

Figure 1: A simple example of some components of a flow

Anomaly detection is the identification of flows which do not conform to an expected pattern or other flows in a dataset. The general process of working and find anomalies with NetFlow includes capturing, sampling, generating, exporting, collecting, analyzing, visualizing and compare [1]

Methods

We installed a set of flow tools on a computer in the UPR **ScienceDMZ**, this refers to a computer subnetwork that is structured to be secure, but without the performance limits that would otherwise result from passing data through a stateful firewall. The Science DMZ is designed to handle high volume data transfers, typical with scientific and high-performance computing, by creating a special DMZ to accommodate those transfers.

One of our flow tools is **SiLK**, we configure SiLK in this subnetwork because it supports the efficient collection, storage, and analysis of network flow data, enabling network security analysts to rapidly query large historical traffic data sets. The other tool is **FlowBAT** is a graphical flow-based analysis tool. We use this tool to detect the anomalies efficiently.

IP Results

We found an anomaly on March 1, 2015. To find what was the anomaly we start searching for all the IP's that was on the network that night.

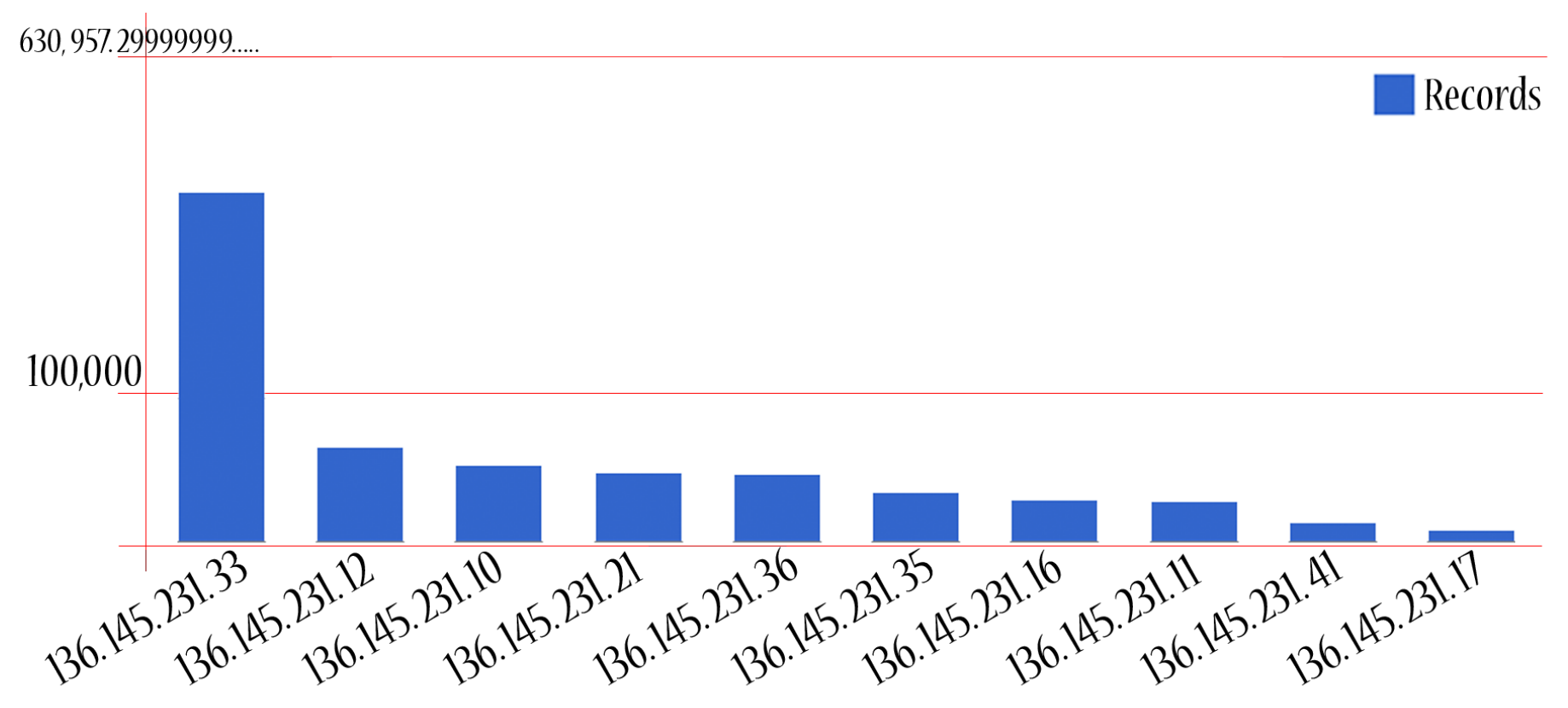


Figure 3: IP's in the network on March 1, 2015

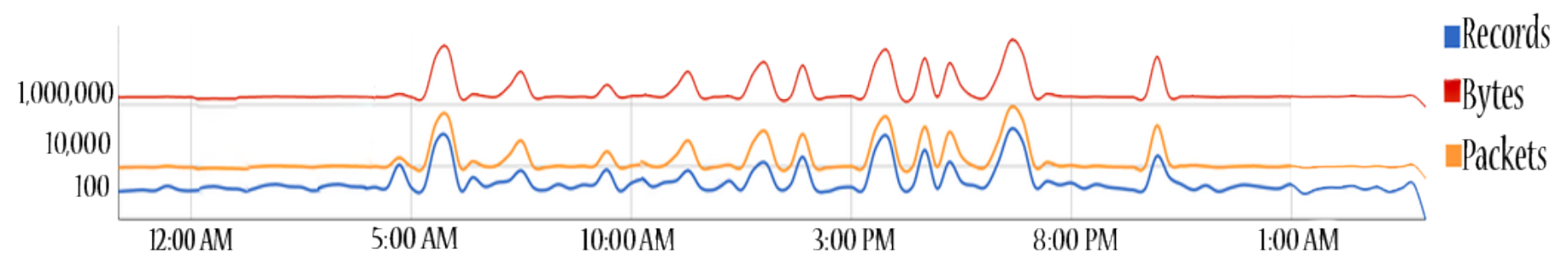


Figure 4: The traffic in March 1, 2015 for the IP 136.145.231.33

Week

With SiLK we collected 168 hours of version 9 flows from the ScienceDMZ, from February 27, 2015 to March 5, 2015.

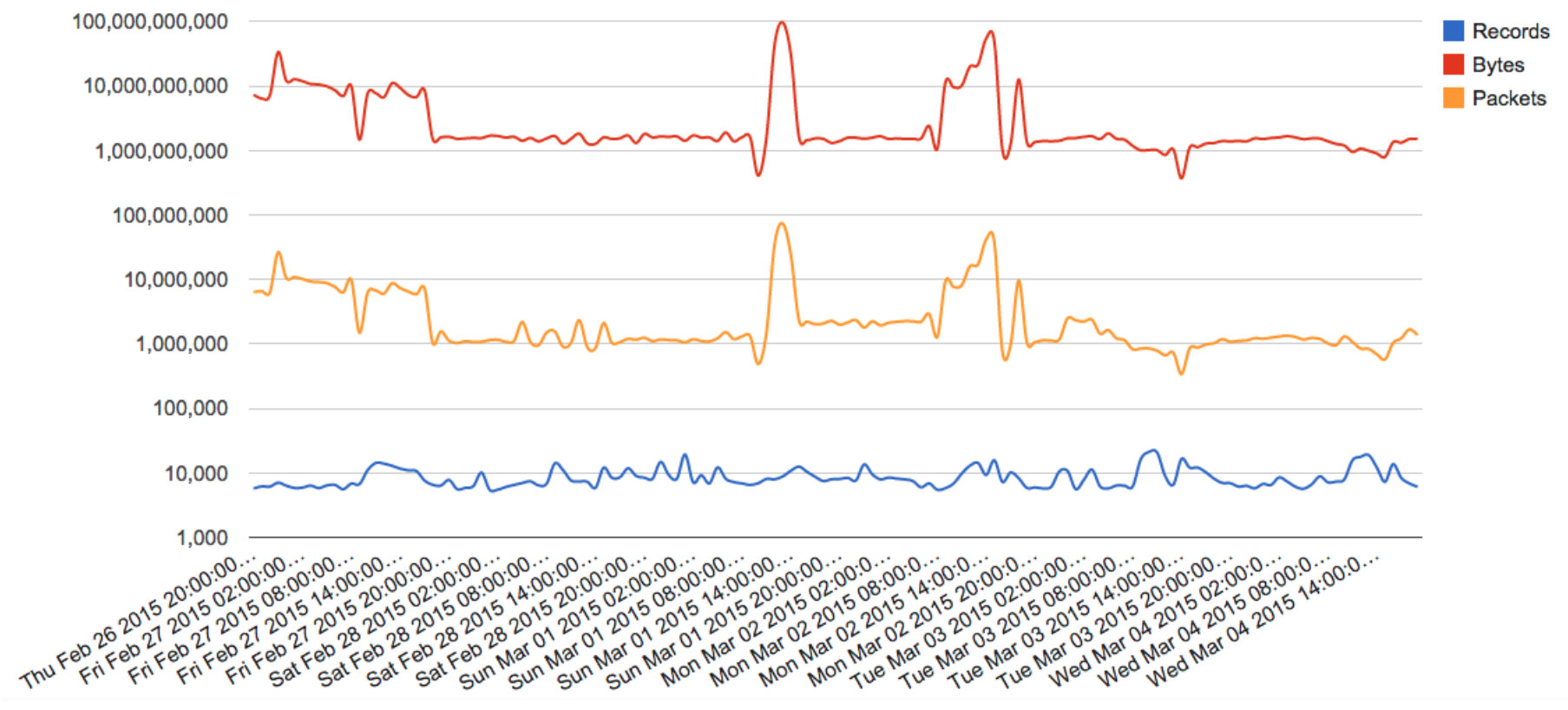


Figure 2: Data Flow count from February 27,2015 to March 5, 2015 in the UPR ScienceDMZ.

We start analyzing the whole week flow data with FlowBAT and find an anomaly on the night of Sunday, March 1, 2015.

Ports Results

Once we identified the IP, we started searching for the port with most activity. We found that was the port 22. And the time with more activity match with our anomaly.

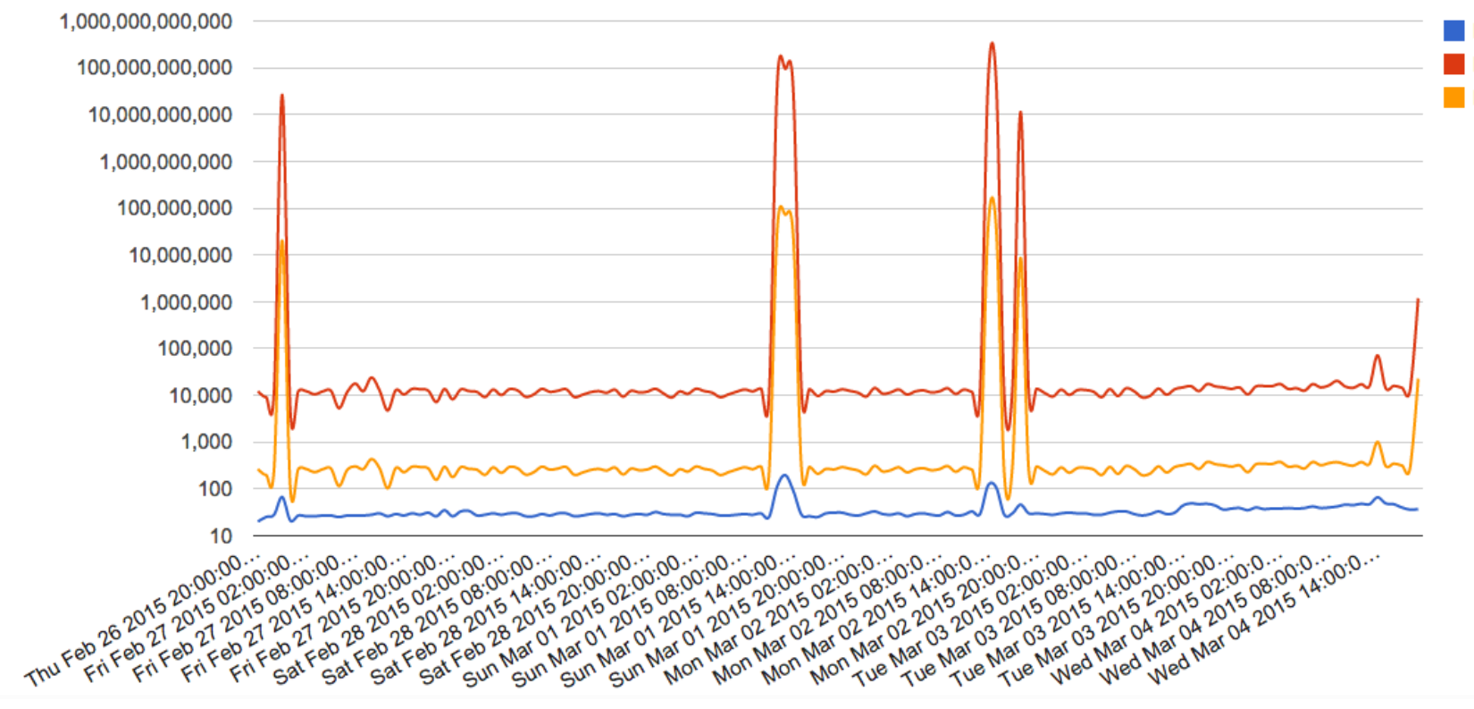


Figure 5: Traffic in port 22 in the week

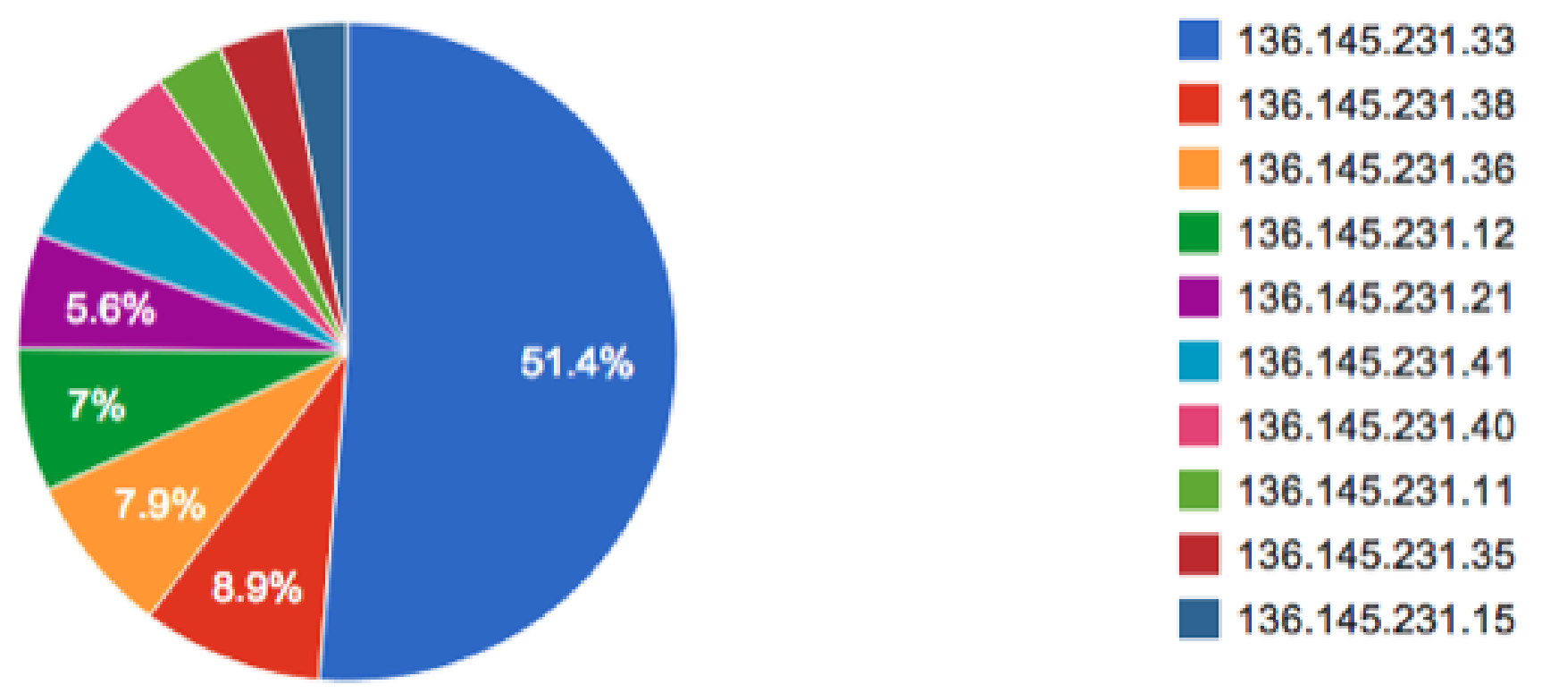


Figure 6: IP's in port 22, on March 1, 2015

Conclusion

Our anomaly was an false positive because that IP is from a professor that collects audio of animals in their natural environment and save that audio nn the server. Our anomaly match with the time of data transfer. We need to collect more data, to learn normal patterns and find anomalies.

Finding a general method for detecting anomalies in flows is hard.

Future Work

In the future, we will explore new approaches to find new techniques. Implement these techniques for anomaly detection to our collection of flows from UPR's network, and compare results with the results of current techniques.

References

[1] B. Li, J. Springer, G. Bebis, and M. H. Gunes.
A survey of network flow applications.
Journal of Network and Computer Applications, 36(2):567–581, 2013.

[2] I. García and H. Ortiz-Zuazaga.
Techniques for anomaly detection in network flows.
`http://ccom.uprrp.edu/~humberto/research/anomaly-detection.pdf`.

[3] R. Bandes, T. Shimeall, M. Heckathorn, and S. Faber.
Using silk for network traffic analysis.
`http://tools.netsa.cert.org/silk/analysis-handbook.pdf`, October 2014.

Acknowledgements

This work is supported by the scholarship Academics and Training for the Advancement of Cybersecurity Knowledge in Puerto Rico (ATACK-PR) suported by the National Science Foundation under Grant No. DUE-1438838.