Improving Satellite Security Through Incremental Anomaly Detection on Large, Static Datasets

Connor Hamlet[1], Matthew Russell[2], Jeremy Straub[3] , Scott Kerlin[3]
[1]Department Computer Science, University of North Carolina at Chapel Hill
[2]School of Engineering and Computational Sciences, Liberty University
[3]Department of Computer Science, University of North Dakota

Overview

Anomaly detection is a widely used technique to detect system intrusions. Anomaly detection in Intrusion Detection and Prevent Systems (IDPS) works by establishing a baseline of normal behavior and classifying points that are at a farther distance away as outliers. The result is an "anomaly score", or how much a point is an outlier. Recent work has been performed which has examined use of anomaly detection in data streams [1]. We propose a new incremental anomaly detection algorithm which is up to 57,000x faster than the non-incremental version while slightly sacrificing the accuracy of results. We conclude that our method is suitable for incremental outlier detection on static datasets on low-resource machines such as satellites.

Background

Several high profile aerospace attacks on low resource machines have occurred in the last decade. In 2008, two NOAA satellites experienced several minutes of interference and a third party achieved all steps to command the spacecraft systems [2]. In 2011, Iran hacked an American RQ-170 and flew it into Iranian controlled airspace [2]. These events highlight the need for anomaly detection on low resource machines. Current anomaly detection methods needed for intrusion detection run in quadratic runtime complexity, which is difficult for low resource machines on large data sets. Algorithms that process data one point at a time update the anomaly scores for each of the point's neighbors, allowing the dataset to grow [3]. Applications with static datasets do not require these points to be updated. Local Outlier Probabilities (LoOP) [4] is a recent static anomaly detection algorithm. LoOP was chosen because the anomaly score is the probability of a point being an outlier in range from zero to one, making our proposed IDPS easier to implement.

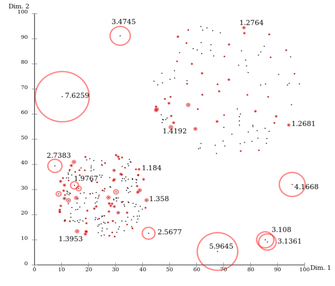

Figure 1. LoOP in Action.

Proposed IDPS and Its Needs

To understand the importance of the proposed algorithm, the IDPS of satellites must be considered. The proposed IDPS detects outliers in satellite sensor data, which may indicate an attack, such as location of the ground station, and signal interference and intensity levels, and other factors which are unlikely to change over time.

Local Outlier Probabilities and Incremental Function

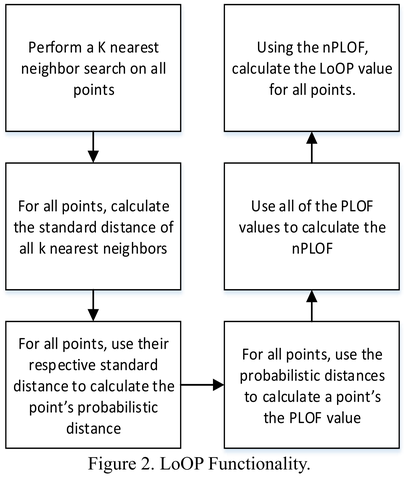

Figure 2. LoOP Functionality.

Implementation

The k-NN search is the most computationally expensive portion of any distance-based anomaly detection algorithm [1]. We implemented appropriate methods to significantly reduce the runtime of the proposed technique.

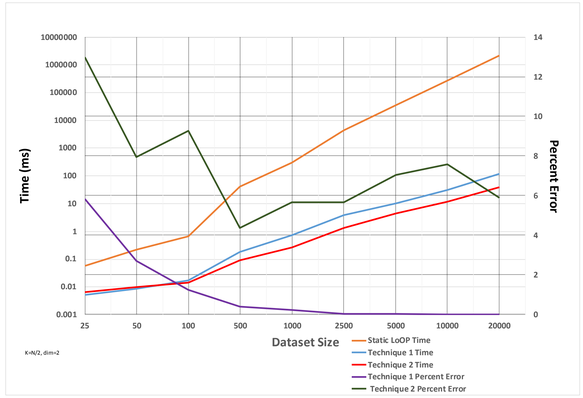

Figure 3. Static Algorithm vs Incremental Modes.

Dataset size (K=N/2)	Exact Static LoOP Time (ms)	Technique 1 Time (ms)	Technique 2 Time (ms)
500	648.5	1.419	0.9362
1000	4879	5.056	2.532
5000	868800	183.3	48.31

Figure 4. Raspberry Pi Tests to Simulate a Low Resource Machine.

$$\sigma(o,S) = \sqrt{\frac{\sum_{s\in S} d(o,s)^2}{|S|}} \qquad \mathrm{PLOF}_S(o) = \frac{\mathrm{pdist}(o,S(o))}{E_{s\in S(o)}[\mathrm{pdist}(s,S(s))]} - 1$$

$$\mathrm{pdist}(o,S) = \sigma(o,S) \qquad \mathrm{nPLOF} = \sqrt{E[(\mathrm{PLOF})^2]}$$

$$\mathrm{LoOP}_S(o) = \max\left\{0, \mathrm{erf}\left(\frac{\mathrm{PLOF}_S(o)}{\mathrm{nPLOF}\cdot\sqrt{2}}\right)\right\}$$

Figure 3. LoOP Equations.

Data set Size (K=N/2)	Technique 1 Time (ms)	%Error Exact	Technique 2 Time (ms)	% Error Approx
100	0.0167	1.23%	0.0139	9.25%
500	0.1805	0.397%	0.0889	4.37%
1000	0.7284	0.22%	0.2607	5.68%
2500	3.868	0.0284%	1.345	5.68%
5000	10.03	0.0268%	4.404	7.03%
10000	30.16	0.00717%	11.87	7.56%
20000	113.1	0.00402%	37.23	5.88%

Figure 5. Incremental Algorithm and % Error from Correct Results.

Results

Comparisons between incremental LoOP and other incremental anomaly detection algorithms show that the proposed techniques are able to perform less operations and possess a faster run time while incurring a slight amount of error. We conclude that this incremental anomaly detection scheme is best suited for low-resource machines that perform outlier detection on data streams with large and unchanging training data.

Future Work

It would be interesting to test this method in higher dimensions with an approximate nearest neighbor search such as Local Sensitive Hashing.

Acknowledgements

This research was funded by the U.S. National Science Foundation (NSF Award #1359224) with support from the U.S. Department of Defense.

References

[1]Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly detection: A survey. *ACM Comput. Surv.* 41, 3, Article 15 (July 2009), 58 pages. DOI=10.1145/1541880.1541882

[2]Fritz, J. *Satellite hacking: A guide for the perplexed. Culture Mandala: The Bulletin of the Centre for East-West Cultural and Economic Studies 10(1), 3, 2013*

[3]D. Pokrajac, N. Reljin, N. Pejcic and A. Lazarevic, "Incremental Connectivity-Based Outlier Factor Algorithm", in *BCS International Academic Conference, London*, 2008, pp. 211-223.

[4]Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. 2009. LoOP: local outlier probabilities. In *Proceedings of the 18th ACM conference on Information and knowledge management* (CIKM '09). ACM, New York, NY, 1649-1652. DOI=10.1145/1645953.1646195