**Implementation of GAN for intrusion detection in large scale Networks**

**Introduction**

To have a secure connection is a very difficult thing in day-to-day life having connected to network of networks. There are many algorithms build to detect the malicious packets but there is still a chance by sending the the packet and waiting for the vulnerable situation to attack and get the required details. So, no algorithm assures to connect with a cent percent secure connection. A different types of Network Anomalies such as DoS, probe, U2R, R2L attacks occur in network. It is proposed an anomaly detection system that monitors the network in three different levels. In the first one, data is collected from simple network management protocol (SNMP) objects and compared to profiles of normal traffic, in order to detect behavior changes. Second level of analysis includes a dependency graph that represents the relationships between SNMP objects. It is used to analyze first level alerts, confirming the occurrence of anomalies in device level. In the third level of analysis, second level alerts are grouped according to network topology information, and network administrators are informed about the context where the anomaly occurred. Also because of APT(Advanced persistent threats) and different kinds of Malwares such as virus, spyware, Trojan horse, the end system in a network can get attacked.

So, with respect to the APTs our model tries to apply GAN based algorithm.

**Dataset Description**

[**Comprehensive, Multi-Source Cyber-Security Events**](http://dx.doi.org/10.17021/1179829),   
Alexander D. Kent. Dataset 2014

This data set represents 58 consecutive days of de-identified event data collected from five sources within Los Alamos National Laboratory’s corporate, internal computer network especially from individual Windows-based desktop computers, servers.

|  |  |
| --- | --- |
| File name | Description |
| auth.txt | authentication events collected |
| proc.txt | process start and stop events collected |
| flows.txt | network flow events collected from central routers |
| dns.txt | Domain Name Service (DNS) lookup events collected from the central DNS servers |
| redteam.txt | specific events taken from the authentication data that present known redteam compromise events |

**Literature Survey**

For the generative model, we have different type of GANs which can be implemented. Among these, a simple JSD based objective function would result be efficient according to mathematics reasoning

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Authors | Architecture Name | Key Take-Away | Objective Function | Applications |  |
| [Ian J. Goodfellow](https://arxiv.org/search/stat?searchtype=author&query=Goodfellow,+I+J), [Jean Pouget-Abadie](https://arxiv.org/search/stat?searchtype=author&query=Pouget-Abadie,+J), [Mehdi Mirza](https://arxiv.org/search/stat?searchtype=author&query=Mirza,+M), [Bing Xu](https://arxiv.org/search/stat?searchtype=author&query=Xu,+B), [David Warde-Farley](https://arxiv.org/search/stat?searchtype=author&query=Warde-Farley,+D), [Sherjil Ozair](https://arxiv.org/search/stat?searchtype=author&query=Ozair,+S), [Aaron Courville](https://arxiv.org/search/stat?searchtype=author&query=Courville,+A), [Yoshua Bengio](https://arxiv.org/search/stat?searchtype=author&query=Bengio,+Y) | GAN |  | JSD divergence |  |  |
| [Jeff Donahue](https://arxiv.org/search/cs?searchtype=author&query=Donahue,+J), [Philipp Krähenbühl](https://arxiv.org/search/cs?searchtype=author&query=Krähenbühl,+P), [Trevor Darrell](https://arxiv.org/search/cs?searchtype=author&query=Darrell,+T) | BiGan |  | JSD divergence |  |  |
|  | Vanilla GAN |  |  |  |  |
|  | DCGAN (Deep Convolutional GAN) |  |  |  |  |
|  | CGAN (Conditional GAN) |  |  |  |  |
| [Martin Arjovsky](https://arxiv.org/search/stat?searchtype=author&query=Arjovsky,+M), [Soumith Chintala](https://arxiv.org/search/stat?searchtype=author&query=Chintala,+S), [Léon Bottou](https://arxiv.org/search/stat?searchtype=author&query=Bottou,+L) | WassersteinGAN | EM distance objective | EM distance |  |  |
| [Ishaan Gulrajani](https://arxiv.org/search/cs?searchtype=author&query=Gulrajani,+I), [Faruk Ahmed](https://arxiv.org/search/cs?searchtype=author&query=Ahmed,+F), [Martin Arjovsky](https://arxiv.org/search/cs?searchtype=author&query=Arjovsky,+M), [Vincent Dumoulin](https://arxiv.org/search/cs?searchtype=author&query=Dumoulin,+V), [Aaron Courville](https://arxiv.org/search/cs?searchtype=author&query=Courville,+A) | Improved WGAN | No weight clipping on WGAN | EM distance |  |  |
| [Xudong Mao](https://arxiv.org/search/cs?searchtype=author&query=Mao,+X), [Qing Li](https://arxiv.org/search/cs?searchtype=author&query=Li,+Q), [Haoran Xie](https://arxiv.org/search/cs?searchtype=author&query=Xie,+H), [Raymond Y.K. Lau](https://arxiv.org/search/cs?searchtype=author&query=Lau,+R+Y), [Zhen Wang](https://arxiv.org/search/cs?searchtype=author&query=Wang,+Z), [Stephen Paul Smolley](https://arxiv.org/search/cs?searchtype=author&query=Smolley,+S+P) | LSGAN(Least Squares GAN) | L2 loss objective | EM distance |  |  |
| [Xin Guo](https://arxiv.org/search/stat?searchtype=author&query=Guo,+X), [Johnny Hong](https://arxiv.org/search/stat?searchtype=author&query=Hong,+J), [Tianyi Lin](https://arxiv.org/search/stat?searchtype=author&query=Lin,+T), [Nan Yang](https://arxiv.org/search/stat?searchtype=author&query=Yang,+N) | RWGAN(Relaxed WGAN ) | Relaxed WGAN framework | EM distance |  |  |
| [Youssef Mroueh](https://arxiv.org/search/cs?searchtype=author&query=Mroueh,+Y), [Tom Sercu](https://arxiv.org/search/cs?searchtype=author&query=Sercu,+T), [Vaibhava Goel](https://arxiv.org/search/cs?searchtype=author&query=Goel,+V) | McGAN(Mean/covariance minimization GAN) | Mean/covariance minimization objective | Mean/covariance minimization objective |  |  |
| [Yujia Li](https://arxiv.org/search/cs?searchtype=author&query=Li,+Y), [Kevin Swersky](https://arxiv.org/search/cs?searchtype=author&query=Swersky,+K), [Richard Zemel](https://arxiv.org/search/cs?searchtype=author&query=Zemel,+R) | GMMN(Generative Moment Matching Networks) | Maximum mean discrepancy objective | Maximum mean discrepancy |  |  |
| [Chun-Liang Li](https://arxiv.org/search/cs?searchtype=author&query=Li,+C), [Wei-Cheng Chang](https://arxiv.org/search/cs?searchtype=author&query=Chang,+W), [Yu Cheng](https://arxiv.org/search/cs?searchtype=author&query=Cheng,+Y), [Yiming Yang](https://arxiv.org/search/cs?searchtype=author&query=Yang,+Y), [Barnabás Póczos](https://arxiv.org/search/cs?searchtype=author&query=Póczos,+B) | MMD GAN  (Maximum mean discrepancy GAN) | Adversarial kernel to GMMN |  |  |  |
| [Marc G. Bellemare](https://arxiv.org/search/cs?searchtype=author&query=Bellemare,+M+G), [Ivo Danihelka](https://arxiv.org/search/cs?searchtype=author&query=Danihelka,+I), [Will Dabney](https://arxiv.org/search/cs?searchtype=author&query=Dabney,+W), [Shakir Mohamed](https://arxiv.org/search/cs?searchtype=author&query=Mohamed,+S), [Balaji Lakshminarayanan](https://arxiv.org/search/cs?searchtype=author&query=Lakshminarayanan,+B), [Stephan Hoyer](https://arxiv.org/search/cs?searchtype=author&query=Hoyer,+S), [Rémi Munos](https://arxiv.org/search/cs?searchtype=author&query=Munos,+R) | Cramer GAN | Cramer distance | Cramer distance |  |  |
| [Youssef Mroueh](https://arxiv.org/search/cs?searchtype=author&query=Mroueh,+Y), [Tom Sercu](https://arxiv.org/search/cs?searchtype=author&query=Sercu,+T) | Fisher GAN | Chi-square objective | Chi-square objective |  |  |
| [Junbo Zhao](https://arxiv.org/search/cs?searchtype=author&query=Zhao,+J), [Michael Mathieu](https://arxiv.org/search/cs?searchtype=author&query=Mathieu,+M), [Yann LeCun](https://arxiv.org/search/cs?searchtype=author&query=LeCun,+Y) | EBGAN([Energy-based G](https://arxiv.org/abs/1609.03126v4)AN) | Autoencoder instead of discriminator |  |  |  |
| [David Berthelot](https://arxiv.org/search/cs?searchtype=author&query=Berthelot,+D), [Thomas Schumm](https://arxiv.org/search/cs?searchtype=author&query=Schumm,+T), [Luke Metz](https://arxiv.org/search/cs?searchtype=author&query=Metz,+L) | BEGAN(B[oundary Equilibrium G](https://arxiv.org/abs/1703.10717)AN | WGAN and EBGAN merged objectives |  |  |  |
| [Ruohan Wang](https://arxiv.org/search/cs?searchtype=author&query=Wang,+R), [Antoine Cully](https://arxiv.org/search/cs?searchtype=author&query=Cully,+A), [Hyung Jin Chang](https://arxiv.org/search/cs?searchtype=author&query=Chang,+H+J), [Yiannis Demiris](https://arxiv.org/search/cs?searchtype=author&query=Demiris,+Y) | MAGAN([Margin Adaptation for G](https://arxiv.org/abs/1704.03817v1)AN | Dynamic margin on hinge loss from EBGAN |  |  |  |