Classifying Computer Processes in the DARPA OpTC dataset

Final Paper for XCS229ii - 003

Andrew Veal
Research Division
UK Government
City State UK
aveal@acm.org

FirstName Surname
Department Name
Institution/University Name
City State Country
email@email.com

FirstName Surname
Department Name
Institution/University Name
City State Country
email@email.com

ABSTRACT – 5 points

The detection of malware and malicious activity in enterprise networks is an ongoing challenge in cybersecurity. Our objective is to build a system that classifies activity associated with a computer process as benign or malicious, using host-based logging data from the computer. By focusing on host activity patterns rather than signatures, we hope to discover behavioral traits that distinguish malicious processes from benign processes. We used supervised learning and 'ground truth' labelled data to build a classification model for the DARPA Operationally Transparent (OpTC) dataset. This is challenging for a number of reasons: the dataset has over 17 billion events; only 0.0016% of events are malicious; the attacks used related *modus operandi*, so classifiers trained on this dataset may not generalize well to other attack scenarios. Our core hypothesis is that we can distinguish between malicious and benign processes using the frequency count of (object, action) events associated with each process as a feature vector. We show that a simple and interpretable baseline model can achieve 88% precision with a recall of 51% on the test set, which suggests our hypothesis is reasonable. We pay special attention to choosing training, validation and test sets from distributions that reflect the data we expect to get in the future. We also show that the DARPA OpTC dataset has the requisite scale, richness and class imbalance to become a new benchmark dataset for cybersecurity researchers.

CCS CONCEPTS

• Intrusion detection • Machine Learning • Big data analytics

KEYWORDS

Cybersecurity dataset, Intrusion detection, Machine Learning

ACM Reference format:

Andrew Veal, FirstName Surname and FirstName Surname. 2021. Classifying Computer Processes in the DARPA OpTC dataset: Final Paper for XCS229ii - 003. In *Proceedings of the 14th ACM Workshop on Artificial Intelligence and Security (AISec 2021). ACM, New York, NY, USA, 8 pages.* https://doi.org/10.1145/1234567890

AISec'2021, November 13, 2021, Seoul, South Korea

978-1-4503-0000-0/18/06...$15.00
https://doi.org/10.1145/1234567890

INTRODUCTION – 10 points

The past twelve months has seen a spate of high-profile cyber-attacks on government, commercial and critical national infrastructure []. As the New York Times reported on the SolarWinds hack []: "Those behind the widespread intrusion into government and corporate networks exploited seams in U.S. defenses and gave away nothing to American monitoring of their systems." The detection of malware and malicious activity in enterprise networks is an ongoing challenge in cybersecurity – state and non-state hackers develop new attacks faster than IT Security teams can deploy signature-based methods to detect new malicious Tactics, Techniques and Procedures (TTP).

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{1}$$

Continuation part of Paragraph Text The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

The **DisplayFormulaUnnum** style is applied only in case of an unnumbered equation. An unnumbered display equation never

contains an equation number to its right, and this unique property distinguishes it from a numbered equation.

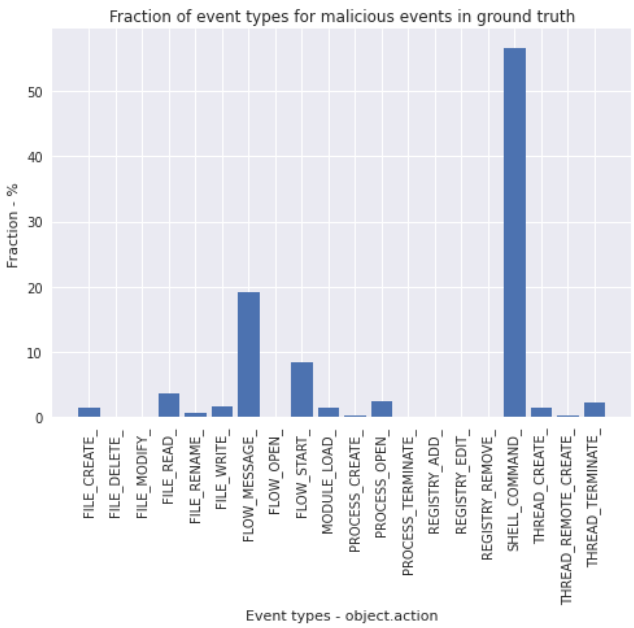

Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

Theorem/Proof/Lemma. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement.

> ....Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract.

Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{1}$$

Continuation part of Paragraph Text The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

1 Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in

1.1 Heading Level 2

In the below paragraph, it is explained how alt-txt value is placed in **MS Word 2010**. To add alternative text to a picture in Word 2010, follow these steps:

1. In a Word 2010 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. Select the **Alt Txt** option from the left-side panel options.
4. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

Below are steps to place alt-txt value in **MS Word 2013/2016**. To add alternative text to a picture in Word 2013/2016, follow these steps:

1. In a Word 2013/2016 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. In the settings at the right side of the window, click on the "Layout & Properties" icon (3rd option).
4. Expand **Alt Txt** option.
5. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

1.1.1 Heading Level 3. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

1.1.1.1 Heading Level 4. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph

text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

1 Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{1}$$

Continuation part of Paragraph Text The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

The **DisplayFormulaUnnum** style is applied only in case of an unnumbered equation. An unnumbered display equation never contains an equation number to its right, and this unique property distinguishes it from a numbered equation.

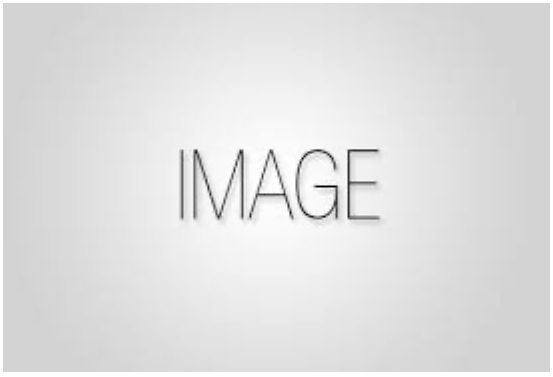

Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

Theorem/Proof/Lemma. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement.

> ....Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract.

Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{1}$$

Continuation part of Paragraph Text The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

1 Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{1}$$

Continuation part of Paragraph Text The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

The **DisplayFormulaUnnum** style is applied only in case of an unnumbered equation. An unnumbered display equation never contains an equation number to its right, and this unique property distinguishes it from a numbered equation.

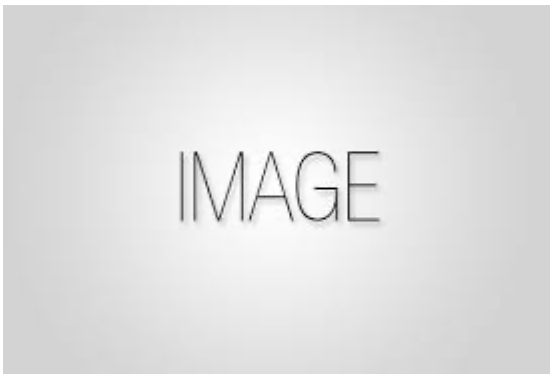

Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

Theorem/Proof/Lemma. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement.

> ....Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract.

Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{1}$$

Continuation part of Paragraph Text The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**

$$\sqrt{b^2-4ac}\,\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

1 Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in

1.1 Heading Level 2

In the below paragraph, it is explained how alt-txt value is placed in **MS Word 2010**. To add alternative text to a picture in Word 2010, follow these steps:

5. In a Word 2010 document, insert a picture.
6. Right click on the inserted picture and select the **Format Picture** option.
7. Select the **Alt Txt** option from the left-side panel options.
8. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

Below are steps to place alt-txt value in **MS Word 2013/2016**. To add alternative text to a picture in Word 2013/2016, follow these steps:

6. In a Word 2013/2016 document, insert a picture.
7. Right click on the inserted picture and select the **Format Picture** option.
8. In the settings at the right side of the window, click on the "Layout & Properties" icon (3rd option).
9. Expand **Alt Txt** option.
10. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

1.1.1 Heading Level 3. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

1.1.1.1 Heading Level 4. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert

paragraph text here. Insert paragraph text here. Insert paragraph text here.

1 Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in

1.1 Heading Level 2

In the below paragraph, it is explained how alt-txt value is placed in **MS Word 2010**. To add alternative text to a picture in Word 2010, follow these steps:

9. In a Word 2010 document, insert a picture.
10. Right click on the inserted picture and select the **Format Picture** option.
11. Select the **Alt Txt** option from the left-side panel options.
12. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

Below are steps to place alt-txt value in **MS Word 2013/2016**. To add alternative text to a picture in Word 2013/2016, follow these steps:

11. In a Word 2013/2016 document, insert a picture.
12. Right click on the inserted picture and select the **Format Picture** option.
13. In the settings at the right side of the window, click on the "Layout & Properties" icon (3rd option).
14. Expand **Alt Txt** option.
15. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

1.1.1 Heading Level 3. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

1.1.1.1 Heading Level 4. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert

paragraph text here. Insert paragraph text here. Insert paragraph text here.

1 Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in

1 Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in

1 Insert Heading Level 1

The updated template, user manuals, samples, and required fonts, all are available at the URL https://www.acm.org/publications/proceedings-template. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in

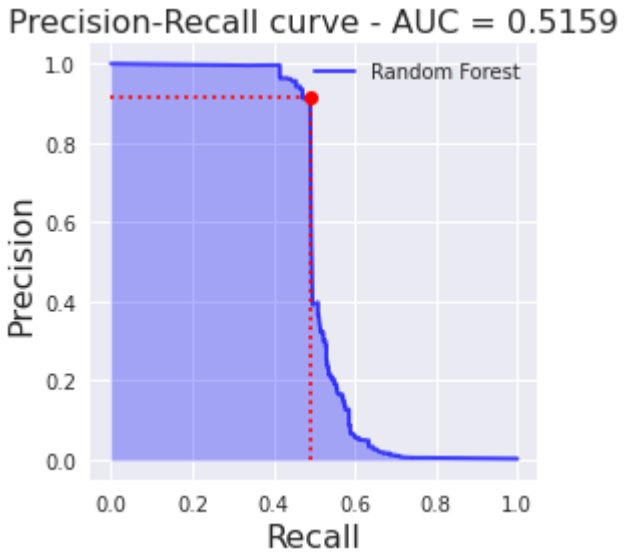

Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

ACKNOWLEDGMENTS

My employer enabled me to access IEEE publications during the Literature Review phase of the project. My employer also allowed me to use our corporate AWS account to do the Extract Transform and Load (ETL) pipeline to reduce the full DARPA dataset down to a scale at which the Machine Learning (ML) project could begin. We did experiments on the ML dataset on Amazon SageMaker – but the ML dataset can be processed on a laptop with Anaconda installed.

REFERENCES – 3 points

[1] Md. Monowar Anjum, Shahrear Iqbal and Benoit Hamelin. 2021. Analysing the Usefulness of the DARPA OpTC Dataset in Cyber Threat Detection Research. arXiv:2103.03080v2. Retrieved from https://arxiv.org/abs/2103.03080 Accepted for *ACM Symposium on Access Control Models and Technologies (SACMAT)*, 16-18 June, 2021, Barcelona, Spain [virtual event]. ACM Inc., New York, NY. DOI: https://doi.org/10.1145/3450569.3463573

[2] BBC. 2010. Profile: Russia's SVR intelligence agency. (June 29, 2010) Retrieved May 12 2021 from https://www.bbc.co.uk/news/10447308

[3] Robert A. Bridges, Tarrah R. Glass-Vanderlan, Michael D. Iannacone, Maria S. Vincent, and Qian (Guenevere) Chen. 2019. A Survey of Intrusion Detection Systems Leveraging Host Data. *ACM Comput. Surv.* 52, 6, Article 128 (November 2019), 35 pages. DOI: https://doi.org/10.1145/3344382

[4] Ben Buchanan. 2020. *The Hacker and the State: Cyber Attacks andthe New Normal of Geopolitics.* Harvard University Press, Cambridge, MA

[5] Anthony Burke. 2020. Robust artificial intelligence for active cyber defence. The Alan Turing Institute. Defence and Security Programme. (March 2020) Retrieved May 12, 2021 from https://www.turing.ac.uk/research/publications/robust-artificial-intelligence-active-cyber-defence

[6] Richard A. Clarke and Robert K. Knake. 2019. *The Fifth Domain: Defending Our Country, Our Companies, and Ourselves in the Age of Cyber Threats.* Penguin Press, New York, NY.

[7] Thomas Cochrane, Peter Foster, Varun Chhabra, Maud Lemercier, Terry Lyons and Cristopher Salvi. 2021. SK-Tree: a systematic malware detection algorithm on streaming trees via the signature kernel. arXiv:2102.07904v3. Retrieved from https://arxiv.org/abs/2102.07904

[8] DARPA. 2020. Operationally Transparent Cyber (OpTC) Data Release. README. Retrieved from http://github.com/FiveDirections/OpTC-data

[9] DARPA. 2020. Operationally Transparent Cyber (OpTC) Data Release. Retrieved from https://drive.google.com/drive/u/0/folders/1n3kkS3KR31KUegn42yk3-e6JkZvf0Caa

[10] DARPA. 2020. OpTC Red Team Ground Truth. Retrieved April 7, 2021 from https://github.com/FiveDirections/OpTC-data/blob/master/OpTCRedTeamGroundTruth.pdf

[11] Neil Daswani and Moudy Elbayadi. 2021. Big Breaches: Cybersecurity Lessons for Everyone. (1st Edition). Apress, Berkeley, CA. DOI: https://doi.org/10.1007/978-1-4842-6655-7

[12] Wei Dong, Charikar Moses, and Kai Li. 2011. Efficient k-nearest neighbor graph construction for generic similarity measures. In *Proceedings of the 20th international conference on World wide web (WWW)*. ACM

[13] GitHub. PyNNDescent. Retrieved from https://github.com/lmcinnes/pynndescent

[14] Sue Halpern. 2021. After the SolarWinds Hack, we have no idea what Cyber dangers we face. *The New Yorker – Daily Comment* (January 25, 2021) Retrieved January 26, 2021 from https://www.newyorker.com/news/daily-comment/after-the-solarwinds-hack-we-have-no-idea-what-cyber-dangers-we-face

[15] David J. Hand. 2009. Mismatched Models, Wrong Results, and Dreadful Decisions. Keynote at *15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, June, 2009 Paris, France recorded June 2009, published September 14, 2009 http://videolectures.net/kdd09_hand_mmwrdd/ (video) http://videolectures.net/site/normal_dl/tag=45840/kdd09_hand_mmwrdd_01.pdf (slides)

[16] David J. Hand. 2009. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Mach Learn*, 77 (2009), 103–123. DOI: https://doi.org/10.1007/s10994-009-5119-5

[17] Robert Hannigan. 2020. SolarWinds hack exploited weaknesses we continue to tolerate. *The Financial Times (December 20, 2020).* Retrieved January 12, 2021 from https://www.ft.com/content/2bed3013-b21f-4b2c-8572-b2da016d1b4e

[18] Haibo He and Edwardo A. Garcia. 2009. Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, Vol.21, No. 9, (September 2009), 1263-1284. DOI: https://doi.org/10.1109/TKDE.2008.239

[19] Max Heinemeyer. 2021. Dissecting the SolarWinds hack without the use of signatures. Retrieved from https://www.darktrace.com/en/blog/dissecting-the-solar-winds-hack-without-the-use-of-signatures/

[20] LANL. 2017. Unified Host and Network Data Set. Retrieved from https://csr.lanl.gov/data/2017/

[21] Guillaume Lemaitre, Fernando Nogueira and Christos K. Aridas. 2016. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research,* 7 (2016) 1-5 DOI: https://dl.acm.org/doi/pdf/10.5555/3122009.3122026

[22] Bartosz Krawczyk. 2016. Learning from imbalanced data: open challenges and future directions. *Prog. Artif. Intell.* 5 (2016) 221-232. DOI: https://doi.org/10.1007/s13748-016-0094-0

[23] MIT Technology Review Insights. 2021. Preparing for AI-enabled cyberattacks. Retrieved from https://www.technologyreview.com/2021/04/08/1021696/preparing-for-ai-enabled-cyberattacks/

[24] National Cyber Security Centre. 2020. Host Based Capability. (May 4, 2020) Retrieved May 11, 2021 from https://www.ncsc.gov.uk/pdfs/information/host-based-capability.pdf

[25] National Cyber Security Centre. 2020. Introducing Host Based Capability (HBC). (November 6, 2020) Retrieved May 11, 2021 from https://www.ncsc.gov.uk/pdfs/blog-post/introducing-host-based-capability-hbc.pdf

[26] Andrew Ng. 2018. Machine Learning Yearning: Technical Strategy for AI Engineers, In the Era of Deep Learning. Draft Version. Retrieved May 6, 2021 from https://www.deeplearning.ai/programs/

[27] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research.* 12(85) (2011) 2825−2830. DOI: https://dl.acm.org/doi/10.5555/1953048.2078195

[28] Nicole Perlroth. 2021. *This is how they tell me the World ends: The Cyber Weapons Arms Race.* Bloomsbury Publishing, London.

[29] Matilda Rhode, Pete Burnap and Kevin Jones. 2018. Early-stage malware prediction using recurrent neural networks. *Computers & Security,* 77 (August 2018), 578-594. DOI: https://doi.org/10.1016/j.cose.2018.05.010

[30] David E. Sanger. 2018. *The Perfect Weapon: War, Sabotage, and Fear in the Cyber Age.* Scribe, London

[31] David E. Sanger, Nicole Perlroth and Julian E. Barnes. 2021. As Understanding of Russian Hacking Grows, So Does Alarm. *The New York* (January 2, 2021) Retrieved January 5, 2021 from https://www.nytimes.com/2021/01/02/us/politics/russian-hacking-government.html

[32] David E. Sanger and Nicole Perlroth. 2021. Russia Appears to Carry Out Hack Through System Used by U.S. Aid Agency. *The New York Times* (May 28, 2021) Retrieved May 28, 2021 from https://www.nytimes.com/2021/05/28/us/politics/russia-hack-usaid.html

[33] Bruce Schneier. 2018. *Click Here to Kill Everybody: Security and Survival in a Hyper-Connected World.* W.W.Norton & Company. New York, NY.

[34] Melissa J. M. Turcotte, Alexander D. Kent and Curtis Hash. 2018. Unified Host and Network Data Set. In *Data Science for Cyber-Security*, Chapter 1 (November 2018), 1-22. World Scientific DOI: https://doi.org/10.1142/9781786345646_001

[35] Aaron Walker and Shamik Sengupta. 2019. Insights into Malware Detection via Behavioral Frequency Analysis using Machine Learning. In *Proceedings of the 2019 IEEE Military Communications Conference (MILCOM)*, 12-14 November, 2019, Norfolk, VA, USA. IEEE *Explore,* 1-6. https://doi.org/10.1109/MILCOM47813.2019.9021034

[36] Charles Wheelus, Elias Bou-Harb and Xingquan Zhu. 2018. Tackling Class Imbalance in Cyber Security Datasets. In *Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI)*, 6-9 July, 2018, Salt Lake City, UT, USA. IEEE *Xplore*, 229-232. https://doi.org/10.1109/IRI.2018.00041