Undergraduate Thesis Prospectus

# BLUESPAWN: An active defense and endpoint detection and response tool

(technical research project in Computer Science)

# Disinformation in the Age of Information

(sociotechnical research project)

by

James McDowell

November 2, 2020

technical project collaborators:

Jacob Smith
William Mayes
Calvin Krist
Kelvin Peng
Dillon Korman
Grant Matteo
Aaron Gdanski
David Smith

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.
*James McDowell*

*Technical advisor*: Yongwhi Kwon, Department of Computer Science

*STS advisor*: Peter Norton, Department of Engineering and Society

**General research problem**

*How can Internet technology abuse be prevented?*

The Internet and associated technologies have forever changed the world and heralded in today's Age of Information. But as with every technological advancement, bad actors seek to find ways to use humanity's new capabilities to cause harm. Cyber crimes are now estimated to cost businesses, individuals, and nations $6 trillion in 2021 alone (Morgan, 2020). Some bad actors' methods are predominantly technical: they find and exploit flaws in software or using software to steal information or interfere with normal operations. Others' are predominantly social: they use technology and apply psychology to influence others via information disorder. This can have widespread impacts on people's beliefs, which while perhaps less directly impactful than information theft or disruptions to operations, has been used to impact large geopolitical events like elections (Select Committee, 2020).

**BLUESPAWN: An active defense and endpoint detection and response tool**

*How can incident response and threat hunting be automated and combined to create a new type of system security tool?*

BLUESPAWN is an endpoint security system used for immediate response and active defense during a cyber attack. Jake Smith, Calvin Krist, Will Mayes, and I worked on this as our capstone research under Professor Kwon of the Computer Science department. Recently, there have been more contributors, including Grant Matteo, Kelvin Peng, David Smith, Aaron Gdanski, and Dillon Korman. Our tool can be found at https://github.com/ION28/BLUESPAWN.

Malware, often in the form of spyware or ransomware, is running rampant. Billions upon billions of dollars are lost from every cyber attack, be it from ransoms, unavailability, or fallout from personally identifiable information (PII) or intellectual property (IP) theft. Nation states threaten elections and critical infrastructure, cyber crime groups destroy hospitals, state sponsored hackers destroy businesses. And it's only getting worse; between 2009 and 2018, the rate of malware infections has risen by an average of 59 percent a year (Purplesec, 2020), and the associated costs for defending is expected to rise by 8.12 percent a year over the next six years (Columbus, 2020). With the stakes so high and the problem so widespread, it's clear that solutions are needed.

Solutions exist for different aspects of the problem, but few are complete or without their flaws. In the case of malware detection, identification, and response, solutions known as antivirus tools or endpoint detection and response tools are the current industry standard. However, these tools generally are limited by a few constraints. First, being widely deployed requires them to be more selective in their findings, meaning they are often less sensitive. They exist primarily as reactionary tools, stopping malware once it's already too late. They exist primarily as static tools that remain on a system, not allowing an analyst to use them to more effectively respond to an incident beyond providing information. Finally, while they do log activities on a system, the brunt of the work in determining the effect and goal of the malware falls to analysts or other tools.

The goal is for BLUESPAWN to address these shortcomings. BLUESPAWN is intended to value sensitivity over selectiving, identifying everything that may be bad, even if it means falsely identifying benign artifacts. It will also be able to proactively apply security settings and mitigate vulnerabilities upon deployment. In addition to running in the background, an analyst

will be able to use it along with their knowledge and skills to more effectively hunt and destroy malware. Finally, the most valuable capability BLUESPAWN will feature is the ability to integrate past logs with what it knows about the malware to identify a network of events related to the malware, making it easier for an analyst to piece together the anatomy of an attack. Almost all defensive security tools are close-sourced, making it challenging to understand how they work or how to make one of your own. BLUESPAWN is made open source in an attempt to help fix this.

BLUESPAWN has received extensive real world testing and evaluation in addition to regular simulated tests. With each set of changes made to the program, a suite of simulated malware samples and malicious indicators are applied to a system, which BLUESPAWN will find, stop, and destroy. However, since this lab condition test may not be indicative of real world effectiveness, BLUESPAWN is also tested in a real time attack scenario, where professional malware developers and industry hackers work to attack a simulated network and deploy custom malware.

Work has been underway on BLUESPAWN since May of 2019, and I have worked on it from the start. In its current form, BLUESPAWN is capable of meeting all of the aforementioned goals to one degree or another, but there is still plenty of room for improvement. By the end of my work on it, the next major component I intend to finish will be the ability to monitor the activities of every program on the system at a very fine grained level, along with the ability to limit the capabilities of any program, stopping malware from ever even being able to try to carry out its malicious task. This will make it easier for defenders to defend, easier for developers to write security tools, and harder for bad actors, which are all important steps towards limiting the impacts of malicious actors.

**Disinformation in the Age of Information**

*How does Russia spread information disorder against America?*

In the age of information, almost everything anyone wants to know is just a Google search away. Unfortunately, so is everything anyone else wants you to know. Information disorder can take several forms: misinformation, spreading falsehoods without the intent to cause harm, disinformation, the malicious spread of falsehoods, and malinformation, the malicious spread of information with context omitted. The recent spread of disinformation by Russia began in earnest with the USSR's 1980 "Operation DENVER" (Nehring and Selvage, 2019), which spread the idea that the USA created AIDS. While this was effective, the advent of Internet technologies has taken information disorder campaigns to a new level; the bar is lower, the audience is larger, and even nation states are falling victim.

Researchers have studied information disorder and the digital systems that propagate it. Dawson and Innes (2019) investigated how Russia built the disinformation campaign it uses against Europe, and Lysenko and Brooks (2018) explored the ways in which Russia is strategically spreading information to disrupt democracies. Not only is disinformation a threat to freedom; it has also been shown by Raman et al. (2020) to be capable of damage to critical infrastructure.

While working towards antithetical goals, Russian and Western intelligence officers and government officials share similar motivating ideas and values in the form of nationalistic pride. Russia's intelligence services are primarily responsible for the spread of Russian information disorder, seeking to undermine Western influence in the world, sow discord in Western nations,

and create geopolitical situations favorable to Russia (Wagner, 2017). While these intelligence officers vary, most distrust Western nations and support the Russian government (Poushter, 2015), and act accordingly. The natural opposition to Russia, then, are the Western nations they seek to undermine, many of which have published reports on the effects of and techniques for combatting Russian information disorder (Department of State, 2020). Nations defending against information disorder have also begun trying to enforce regulations on businesses providing its media (Levush, 2019). Western experts combatting this information disorder also vary, but surely most see themselves as champions of democracy and freedoms, a powerful motivation to defend against Russia's efforts.

Outside the clash of global superpowers, most social groups are primarily interested in limiting the effects of information disorder. A number of nonprofits, such as FirstDraft News, have emerged to sift through information disorder to provide accurate reports (First Draft News, 2020). But such efforts are no substitute for a conscientiously skeptical reader. Confirmation bias leads people to seek out information agreeing with their beliefs, often neglecting to verify it (Wason, 1960), and making them unwitting agents of information disorder (Ciampaglia, 2018). As these inadvertent spreaders are not united behind any singular goal, they are not a social group. That said, they have an important role, and their impact must be considered.

The last category of group involved are those wishing to protect their financial interests. For media companies, information disorder is a high-stakes business hazard. Companies' reputations are at immediate risk; over the long term, other risks include incurring fines, legal liability, or public regulation. To preserve the integrity of their platforms and thereby protect their financial interests, some online media have launched their own information disorder prevention campaigns (Twitter, 2020). This usually involves monitoring ads, detecting bot

accounts, and identifying information disorder campaigns both internally and via working with governments.

## References


Ciampaglia, G. (2018, Jun 21). Biases Make People Vulnerable to Misinformation Spread by Social Media. *Scientific American.* https://www.scientificamerican.com/article/biases-make-people-vulnerable-to-misinformation-spread-by-social-media/

Columbus, L. (2020, Apr 6). 2020 Roundup Of Cybersecurity Forecasts And Market Estimates. https://www.forbes.com/sites/louiscolumbus/2020/04/05/2020-roundup-of-cybersecurity-forecasts-and-market-estimates/

Dawson, A., & Innes, M. (2019, Jun 1). How Russia's Internet Research Agency Built its Disinformation Campaign. *Political Quarterly*, 90(2), 245 - 256.

Department of State (2020, Oct 1). *GEC Special Report: Russia's Pillars of Disinformation and Propaganda - United States Department of State*. https://www.state.gov/russias-pillars-of-disinformation-and-propaganda-report/

First Draft News. (2020, Oct 29). About. Retrieved November 02, 2020, from https://firstdraftnews.org/about/

Levush, R. (2019, Sep 1). Government Responses to Disinformation on Social Media Platforms: Comparative Summary. https://www.loc.gov/law/help/social-media-disinformation/compsum.php

Lysenko, V., & Brooks, C. (2018). Russian information troops, disinformation, and democracy. *First Monday*. doi:10.5210/fm.v22i5.8176

Morgan, S. (2020, Oct 26). Global Cybercrime Damages Predicted To Reach $6 Trillion Annually By 2021. https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021/

Nehring, C., & Selvage, D. (2019, Jul 22). Operation "Denver": KGB and Stasi Disinformation regarding AIDS. https://www.wilsoncenter.org/blog-post/operation-denver-kgb-and-stasi-disinformation-regarding-aids

Poushter, J. (2015, Jun 12). 6 charts showing how Russians see their country and the world. https://www.pewresearch.org/fact-tank/2015/06/12/6-charts-showing-how-russians-see-their-country-and-the-world/

Purplesec. (2020, Oct 3). 2019 Cyber Security Statistics Trends & Data. https://purplesec.us/resources/cyber-security-statistics/

Raman, G., AlShebli, B., Waniek, M., Rahwan, T., & Peng, J. C. (2020, Aug 12). How weaponizing disinformation can bring down a city's power grid. *PLoS ONE*, 15(8), 1 - 14.

Select Committee on Intelligence of the United States Senate. *Russian Active Measures, Campaigns, and Interference in the 2016 U.S. Election Volume 2: Russia's Use of Social Media With Additional Views*. 116th Congress, 2020.

Twitter. (2020, Jun 12). Disclosing networks of state-linked information operations we've removed. https://blog.twitter.com/en_us/topics/company/2020/information-operations-june-2020.html

Wagner, K. (2017, Nov 1). These are some of the tweets and Facebook ads Russia used to try and influence the 2016 presidential election. *Vox.* https://www.vox.com/2017/10/31/16587174/fake-ads-news-propaganda-congress-facebook-twitter-google-tech-hearing

Wason, P. C. (1960). On the Failure to Eliminate Hypotheses in a Conceptual Task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140. https://doi.org/10.1080/17470216008416717