# The Use of AI Models to Combat Hardware-Based Cyber Threats

**Ariel Sischy**
221003350

## 1. Introduction

This paper will explore different hardware vulnerabilities and exploits attackers can use, and the use of AI in combatting such attacks. The following sections will be the main focus of this paper. The use of AI, to attack and train AI models to get a wide variety of attack paths and data for future sensors to detect. The use of AI with trained data models to quickly detect and respond to threats that humans could not detect or react to in time. The detrimental effects attacks on the supply chain can have on those involved, and its use in accessing higher security systems.

1. The Use of AI Models to Combat Hardware-Based Cyber Threats
2. Ariel Sischy, 221003350

### 1.1. Last Update

This research paper was last updated on 09 September 2023.

## 2. Modular Multilevel Converters

Modular Multilevel Converters are a type of voltage source converter that handles medium to high power conversions. The low cost, scalability, high-quality output performance and great modularity have resulted in MMCs being widely used in high voltage direct current transmission systems that enable the transfer of electricity over long distances, medium voltage motor drives, renewable energy systems, battery systems and more (Wang, et al., 2020).

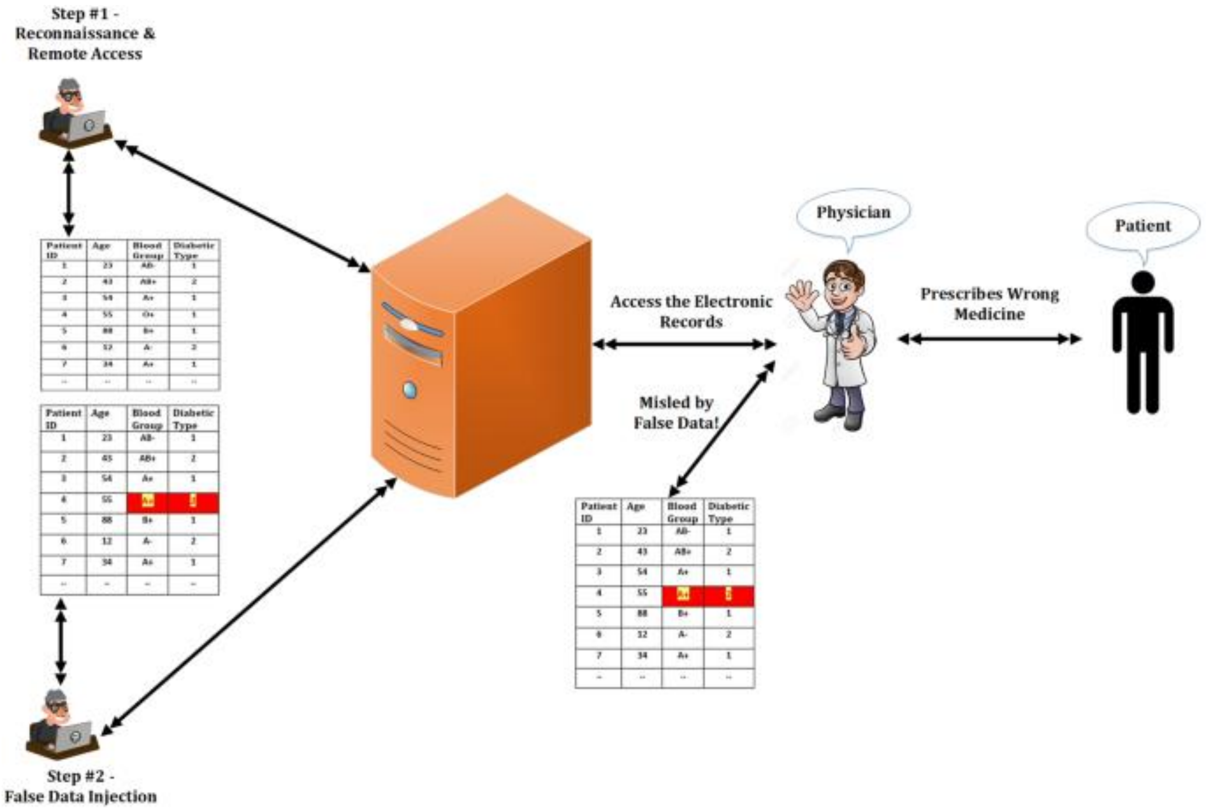
The implementation of an MMC comprises of both a cyber, and physical level, leading to a cyber-physical system (CPS). For the purpose of this article, we will only be looking at the physical level. At the hardware level, the semiconductor components that make up the MMC are switched on and off by the control system. It has been found that almost all research reports point to the CPS always reporting correct information. The vulnerability that is exploited here is the assumption that the data being reported is correct (Burgos-Mellado, et al., 2023).

However, this is not the case when the MMC has been affected by a cyberattack. One such attack is the False Data Injection Attack (FIDA).

## 2.1. False Data Injection Attacks

An FDI attack is where an attacker gains illegal access to the system, allowing them to inject false data. This false data results in the control system incorrectly turning the MMC semiconductors off or on. The CPS will then read and process this false information, leading to an incorrect system state (Ahmed & Pathan, 2020). This incorrect state will result in the SM capacitor voltages not being regulated, leading to high and/or low voltages (Burgos-Mellado, et al., 2023).

Hence, this high/low voltage state, can both damage the health of any battery lifespan, damage electric appliances and trip the M2C protection system (Zhou, Huang, & Pecht, 2018). Such an attack would have devastating effects on both home systems, businesses and power grids. An attack on the power grid could be used for terrorist actions, or criminal purposes such as de-activating a security system, or shutting down a business's service/website as a form of denial of service attack.



| Patient ID | Age | Blood Group | Diabetic Type |
|---|---|---|---|
| 1 | 23 | AB- | 1 |
| 2 | 43 | AB+ | 2 |
| 3 | 54 | A+ | 1 |
| 4 | 55 | O+ | 1 |
| 5 | 88 | B+ | 1 |
| 6 | 12 | A- | 2 |
| 7 | 34 | A+ | 1 |
| -- | -- | -- | -- |

| Patient ID | Age | Blood Group | Diabetic Type |
|---|---|---|---|
| 1 | 23 | AB- | 1 |
| 2 | 45 | AB+ | 2 |
| 3 | 54 | A+ | 1 |
| 4 | 55 | [illegible] | [illegible] |
| 5 | 88 | B+ | 1 |
| 6 | 12 | A- | 2 |
| 7 | 34 | A+ | 1 |
| -- | -- | -- | -- |

| Patient ID | Age | Blood Group | Diabetic Type |
|---|---|---|---|
| 1 | 23 | AB- | 1 |
| 2 | 43 | AB+ | 2 |
| 3 | 54 | A+ | 1 |
| 4 | 55 | [illegible] | [illegible] |
| 5 | 88 | B+ | 1 |
| 6 | 12 | A- | 2 |
| 7 | 34 | A+ | 1 |
| -- | -- | -- | -- |

*Figure 1. Visual of a false data injection attack (Ahmed & Pathan, 2020)*

## 2.2. Mitigation of FDI Attacks

Firstly, FDIA detectors should be put in place to detect such attacks. To further train the detectors, the use of a reinforcement learning (RL)-based method should be used to expose the existing vulnerabilities in the detectors. A RL method generates a multitude of complex attacks meant to bypass the current FDIA detectors. This will highlight the current detector weaknesses that need to be addressed and improved on for future FDIA detector iterations (Burgos-Mellado, et al., 2023).

The use of a RL model instead of a human attack, is crucial due to both the sheer magnitude of different types of attacks, and its ability to test, find and exploit any possible weakness that the detectors might have. Such a task would be both too time consuming, error prone and inefficient for a human to manually do. It is instead better for a human to make use of the data the model uncovers.

In Conclusion, we can make use of such RL models in not only the limited setting of testing FDIA detectors, but any type of system. This is incredibly important and useful for cybersecurity specialists in ensuring that a security system is up to code and has no vulnerabilities (Jaber & Fritsch, 2023).

# 3. Universal Serial Bus

Universal Serial Buses (USBs) are a plug-and-play interface that allows information transfer between two devices (Hope, 2022). This data transfer is a vulnerability that attackers can exploit to inject malicious scripts into the computer device through keypresses (Gurčinas, Dautartas, Janulevičius, Goranin, & Čenys, 2023). Almost every user interactable, non-mobile phone devices, has some sort of USB port, thus Making it paramount to address such a threat.

One of the most common methods of deploying such an attack, is through the use of flash drives. Attackers make use of basic vulnerabilities in human behaviour through social engineering to infect computers with malware. Such attacks include leaving a flash drive on the floor, or parking lot of a company, playing on the curiosity of an employee who may plug the device in their computer. Attackers can also disguise the device as a reward or prize, lowering a person’s guard. In both cases, the malware/scripts can infect not only someone’s personal device, but spread through the business network, infecting every device on it (Gurčinas, Dautartas, Janulevičius, Goranin, & Čenys, 2023).

One of the most common infections is called a keypress attack.

## 3.1. Keypress Attacks

Keypress attacks work by running scripts, usually in the background where the user is unaware, that creates an emulated keyboard. This allows the script to emulate any action or keypress the user is able to accomplish, such as opening up a connection to the attacker’s server. Keypress attacks can either run stealthily in the background, logging whatever the user types to be sent back to the attacker’s server, or actively even opening up a connection for the attacker to inject more malicious software. The scripts are able to accomplish this, as they can send a false keypress to the system. The success of such attacks is in the sheer scope of the number of simultaneous attacks they are making. All it takes is 1 among 1 000 attacks to work for a keypress attack operation to be deemed successful (Identity, 2023)

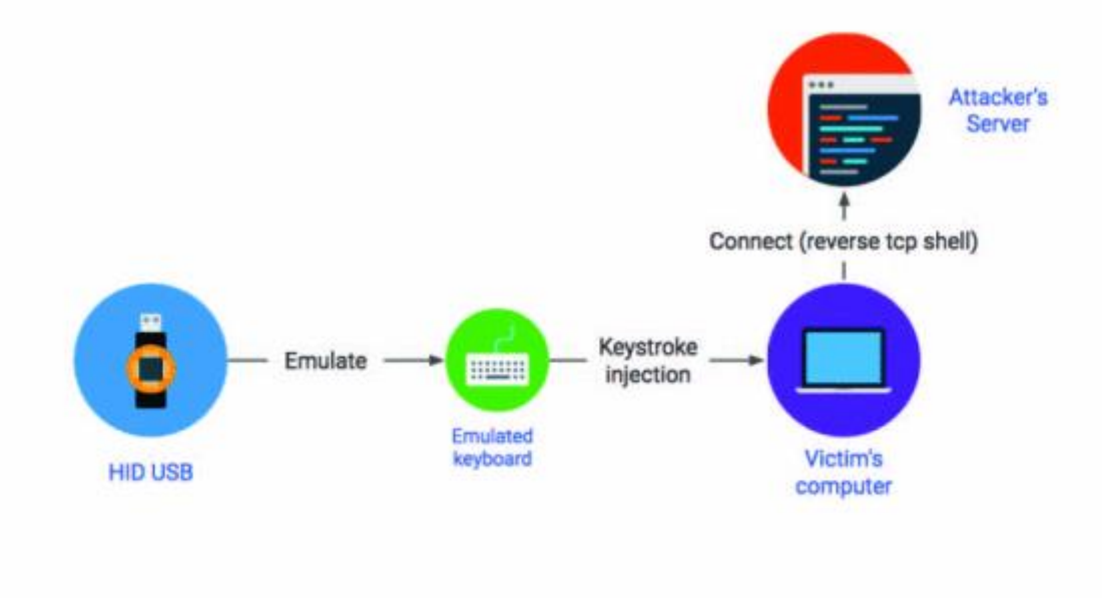


*Figure 2. Visual of an USB keypress injection attack (Lucian, 2016)*

## 3.2. Mitigation of Keypress Attacks

Such an attack happens at the speed of data, making it not possible for the user to counteract such in time. Therefore, a proposed solution of training various machine learning models over different keystroke features, such as typing speed, consistencies in such typing speeds and how long a key is held down for. This trained model will be a real-time, continuous checking and comparing the keystroke features against its own dataset, to determine if the keystrokes are a malicious attack or not (Gurčinas, Dautartas, Janulevičius, Goranin, & Čenys, 2023).

## 3.3. Why the use of AI is Paramount in the Mitigation of Keypress Attacks

Humans require 0.3-3.5 seconds to properly react to stimulus (Aleksander & Pawel, 2021). In conjunction with the onset of panic and delay caused by such, we cannot rely on humans to defend against such cyber attacks. Once an attack has had enough time to gain control of the system, it is able to lock the user out from doing anything. In addition, your everyday office worker is not a highly skilled or cyber security educated individual and would be able to take action against such attacks.

Therefore, it is up to the OS's security systems, and private-business network security to secure the systems for their users and/or customers. AI only requires a dataset, and quickly reacts to threats with machine-speed and the ability to run in the background (Coufalíková, Klaban, & Šlajs, 2021). The advantages of having such a machine-model are its ability to use a trained dataset to almost instantaneously detect such attacks, and put a stop to them before any long term damage or personal information is stolen. Not only that, but an AI model and a trained data set can be copied (sold) to meet the needs of companies with similar security needs.

In addition, it will be able to catch keypress attacks that are not easily visible to the user, such as a script to steal private information, such as banking details and credit card information (Negi, Rathore, & Sadhya, 2014). Such information theft can be financially devastating for the victims involved, who would be none the wiser of how such a disaster occurred.

# 4. Attacks on the Supply chain

Supply chain attacks are when attackers exploit security vulnerabilities in one or more systems of a business to either gain access to that business, or more commonly to gain access to the business that uses or buys the products or services of the affected business (Coufalíková, Klaban, & Šlajs, 2021).

Cyber attacks have the potential to take down a business and its production capabilities for potentially hours or even days at a time (Great American Insurance Company, 2023). It is important to not only look at the effect this will have on that individual business or factory, but also on every other business that is relying on the goods or services produced by such. Such an attack will result in both a loss of time and revenue, but also a loss of trust from every other business that was relying on the affected business and has also now made a loss (Great American Insurance Company, 2023).

In the case of using one business to gain access to another, we can look at “Operation Aurora” as an example. In this case, the attackers exploited the weaker security in lower tier businesses and used these manufacturers as a vulnerability to gain access to high tier defence contractors. This was done to gain access to classified information on the production of parts used by these larger companies in their own production plants (Coufalíková, Klaban, & Šlajs, 2021).

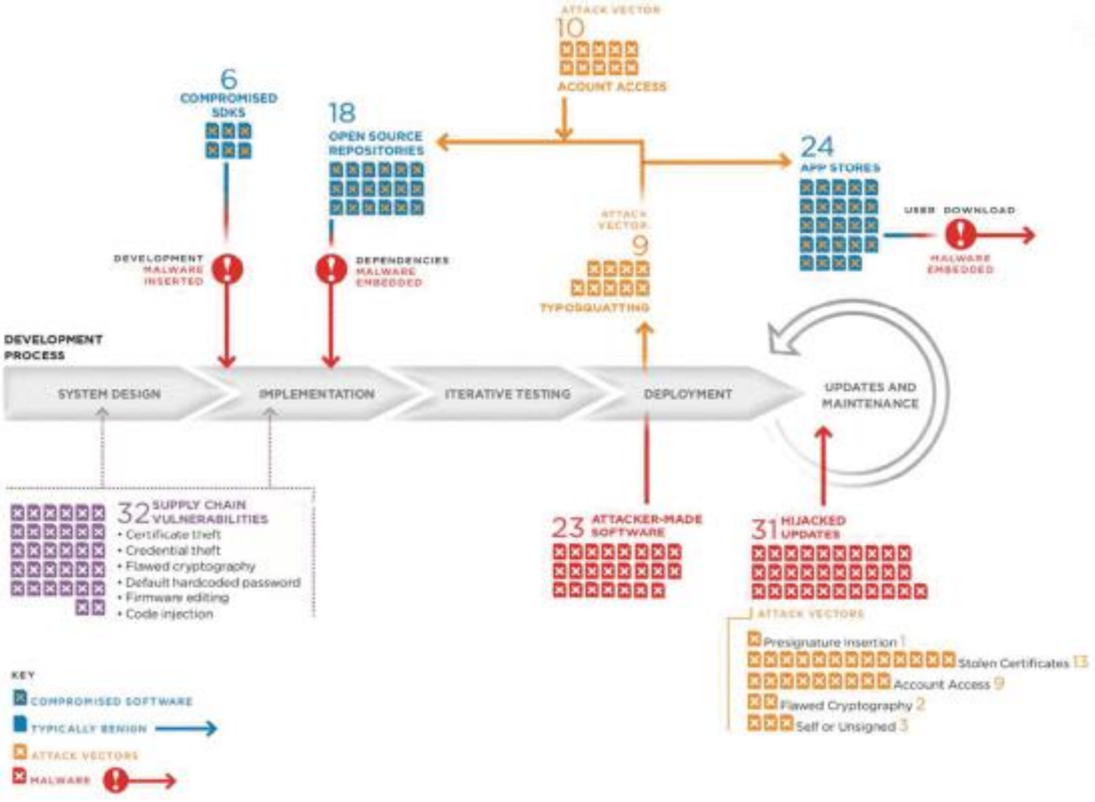


*Figure 3. Examples of Supply Chain Attacks (Coufalíková, Klaban, & Šlajs, 2021)*

## 4.1. Mitigation of Supply Chain Attacks

It is paramount for protection and countermeasures against supply-chain attacks to be put in place. This is especially important for government critical infrastructures such as oil and electricity. The issue with such protection is how complicated implementation is in real world environments, as usual protection software such as antiviruses poses no obstruction to such attacks (Coufalíková, Klaban, & Šlajs, 2021).

To combat and detect these unconventional attacks, more complex approaches are needed, such as the Zero Trust Approach, continuous monitoring of the network, and pre-deployment checks.

### 4.1.1. Zero Trust Approach

This approach is where you do not automatically believe that anything in a set of parameters. Such as user Bob. You can accomplish this by limiting user rights to their roles and or devices, and limiting access by checking their location and the type of connection request being made (Coufalíková, Klaban, & Šlajs, 2021).

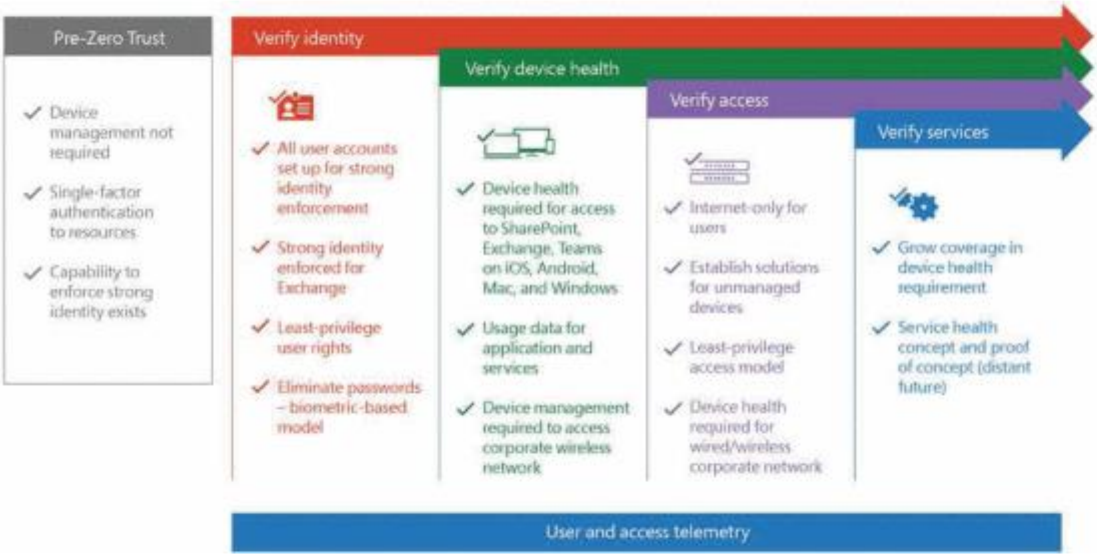

Pre-Zero Trust
✓ Device management not required
✓ Single-factor authentication to resources
✓ Capability to enforce strong identity exists
Verify identity
✓ All user accounts set up for strong identity enforcement
✓ Strong identity enforced for Exchange
✓ Least-privilege user rights
✓ Eliminate passwords – biometric-based model
Verify device health
✓ Device health required for access to SharePoint, Exchange, Teams on iOS, Android, Mac, and Windows
✓ Usage data for application and services
✓ Device management required to access corporate wireless network
Verify access
✓ Internet-only for users
✓ Establish solutions for unmanaged devices
✓ Least-privilege access model
✓ Device health required for wired/wireless corporate network
Verify services
✓ Grow coverage in device health requirement
✓ Service health concept and proof of concept (distant future)
User and access telemetry

*Figure 4. Zero Trust Approach (Coufalíková, Klaban, & Šlajs, 2021)*

### 4.1.2. Continuous Monitoring of the Network

If an attacker has gained access to the businesses network, the use of a behavioural AI that can recognise unusual traffic at machine speed is crucial to catch any anomalies detected by the business to eliminate before any further damage can be done. There are specific indicators that must be updated periodically (this can be done automatically) to maintain the accuracy and responsiveness of the system (Coufalíková, Klaban, & Šlajs, 2021).

### 4.1.3. Pre-Deployment Checks

Before any new code is updated or deployed into production, it is best practice to perform a set of safety checks to prevent any malicious or unauthorized code from making it into the system (Varshney, Joshi, Sardana, Natarajan, & Chaudhuri, 2012).

The advantage of this approach is the use of a testing environment is already used for the testing of bugs and errors, making such an implementation resource and time efficient.

In conclusion, we can therefore see the devastating cascading effects that a cyber attack on a supply chain can have on a multitude of businesses in that chain, and why such attacks must be addressed.

# 5. Conclusion

In conclusion, False Injection Data Attacks and how systems that believe all input data is true, can be exploited. We can see the dangers of USB keypress attacks and the ease and prevalence they are distributed in society. The catastrophic effects and unusual security breaches caused by supply chain attacks, and the new outlook on using non-conventional approaches to combat such attacks. In all of these cases, we have seen how the use of AI models for advanced data model training, detection and mitigation of attacks is crucial to staying up to date in the ever evolving fight against cyber attacks

# Bibliography

Ahmed, M., & Pathan, A. (2020). False data injection attack (FDIA): an overview and new metrics for fair evaluation of its countermeasure. *Complex Adaptive Systems Modeling,* 8. doi:https://doi.org/10.1186/s40294-020-00070-w

Aleksander, R., & Pawel, C. (2021). Human awareness versus Autonomous Vehicles view: comparison of reaction times during emergencies. *2021 IEEE Intelligent Vehicles Symposium (IV),* 732-739. doi:10.1109/IV48863.2021.9575602

Burgos-Mellado, C., Zuñiga-Bauerle, C., Muñoz-Carpintero, D., Arias-Esquivel, Y., Cárdenas-Dobson, R., Dragičević, T., . . . Watson, A. (2023). Reinforcement Learning-Based Method to Exploit Vulnerabilities of False Data Injection Attack Detectors in Modular Multilevel Converters. *IEEE Transactions on Power Electronics, 38*(7), 8907-8921. doi:10.1109/TPEL.2023.3263728

Coufalíková, A., Klaban, I., & Šlajs, T. (2021). Complex strategy against supply chain attacks. *2021 International Conference on Military Technologies (ICMT)*, 1-5.

Great American Insurance Company. (2023). *How Much Can Cyber Attacks Hurt Your Business?* Retrieved from PolicySweet.com: policysweet.com/news/article/how-much-can-cyber-attacks-hurt-your-business#:~:text=So%2C%20a%20denial%20of%20a,falling%20victim%20to%20a%20cyberattack.

Gurčinas, V., Dautartas, J., Janulevičius, J., Goranin, N., & Čenys, A. (2023). A Deep-Learning-Based Approach to Keystroke-Injection Payload Generation. *Electronics, 12*(13), 2894. doi:10.1109/SPIN52536.2021.9566083

Hope, C. (2022, 18 10). *USB.* Retrieved from ComputerHope: https://www.computerhope.com/jargon/u/usb.htm

Identity, B. (2023). *Keystroke Logging*. Retrieved from Beyond Identity: https://www.beyondidentity.com/glossary/keystroke-logging

Jaber, A., & Fritsch, L. (2023). Towards AI-powered Cybersecurity Attack Modeling with Simulation Tools: Review of Attack Simulators. *Barolli, L. (eds) Advances on P2P, Parallel, Grid, Cloud and Internet Computing. 3PGCIC 2022. Lecture Notes in Networks and Systems*, 1. doi:https://doi.org/10.1007/978-3-031-19945-5_25

Lucian, A. (2016, 08 06). *Spreading Malware Through Dropped USB Sticks Could Be Highly Effective, Research Finds*. Retrieved from tom'sHardware: https://www.tomshardware.com/news/dropped-usb-sticks-spreads-malware,32391.html

Negi, A., Rathore, S. S., & Sadhya, D. (2014). Computer forensics in IT audit and credit card fraud investigation - for USB devices. *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*, 730-733. doi:10.1109/IndiaCom.2014.6828058

Varshney, G., Joshi, R. C., Sardana, A., Natarajan, S. N., & Chaudhuri, S. R. (2012). 2012 International Conference on Computer & Information Science (ICCIS). *A Conceptual Framework for Pre Deployment Network Analysis of specification driven systems*, 868-873. doi:10.1109/ICCISci.2012.6297148

Wang, Y., Aksoz, A., Geury, T., Ozturk, S., Kivanc, O., & Hegazy, O. (2020). A Review of Modular Multilevel Converters for Stationary Applications. *Power Electronic Applications in Power and Energy Systems*, 7719. doi:https://doi.org/10.3390/app10217719

Zhou, Y., Huang, M., & Pecht, M. (2018). An Online State of Health Estimation Method for Lithium-ion Batteries Based on Integrated Voltage. *IEEE International Conference on Prognostics and Health Management (ICPHM)*, 1-5. doi:10.1109/ICPHM.2018.8448947