Don't Drink the Cyber: Extrapolating the Possibilities of Oldsmar's Water Treatment Cyberattack

James Cervini, Aviel Rubin and Lanier Watkins
The Johns Hopkins University, Baltimore, USA
jschaf12@jhu.edu
rubin@jhu.cs.edu
lanier.watkins@jhuapl.edu

Abstract: Water treatment represents an essential critical infrastructure sector which has a direct impact on the health and well-being of its customers. Water treatment is often performed by municipalities with very limited budgets for cybersecurity resources. These underfunded, high-impact, targets represent an emerging cyber warfare attack-surface paradigm which poses a direct threat to the quality of life for millions of people. On February 5th, 2021, a water treatment plant in Oldsmar, Florida was the victim of an attempted cyberattack. This attack commanded the system to add a dangerous amount of sodium hydroxide to water which supplied thousands. Direct exposure to sodium hydroxide causes painful burns to the exposed area with permanent internal damage likely upon ingestion. A system operator noticed this malicious behaviour and corrected the situation, minimizing the attack's impact. This paper outlines the attack and illustrates how minor modifications to the attacker's tactics, techniques, and procedures could have resulted in a cyber-derived catastrophe for thousands of unsuspecting citizens. Lastly, this paper explores the effectiveness of various low-cost cyber-physical security technologies when pitted against differing attacker models in these theoretical scenarios. These cybersecurity solutions are evaluated by cost, ease of use, implementation difficulty, and ability to support safe operation continuity when faced with adversary behaviour. The results of this evaluation illuminate a path forward for low-cost threat mitigation which increases the difficulty to compromise these critical cyber-physical systems. With attacks targeting industrial control systems on the rise, the Oldsmar water treatment cyberattack represents more than an individual incident, it can be viewed as a reflection of the current status of thousands of similar critical infrastructure systems that have yet to be caught in crosshairs of a competent and willing adversary with financial incentives and cyber warfare mission requirements serving as impetus for adversary willingness and any resulting large-scale cyber cataclysm.

Keywords: operational technology, water treatment, Oldsmar Florida, cybersecurity, cyber warfare, cyber attack

1. Introduction

The water and wastewater systems sector is labelled as a vital critical infrastructure sector by the United States' Department of Homeland Security (DHS), describing this sector as "essential to modern life" (DHS, 2015). Water treatment responsibilities are commonly performed by municipalities with meagre budgets for cybersecurity resources. According to a report which surveyed water sector utilities, 44.8 percent claimed less than 1 percent of their overall budget was dedicated to operational technology (OT) cybersecurity (ThreatLocker Inc., 2021). The high impact of water treatment disruption and malicious manipulation when paired with an underfunded defensive cyber-posture results in an emerging cyber warfare paradigm which poses a direct threat to the quality of life for millions of people. The 2021 water treatment cyberattack orchestrated in Oldsmar, Florida represents a relevant example of cybersecurity shortcomings prevalent in underfunded municipalities. While Oldsmar is home to approximately 15,000 residents, coordinated cyber warfare campaigns targeting these vulnerable critical systems represent asymmetric threats which could impact millions.

This paper contributes to the domain by researching a relevant OT municipal cyberattack with unrealized impact while framing it in the context of cyber warfare at scale. Furthermore, this paper illustrates how an increased likelihood for impact realization is possible while outlining mitigations mindful to the operational constraints of municipalities. Section two of this paper will describe and dissect this attack, with section three ultimately showing how slight modifications to the attacker's tactics, techniques, and procedures (TTPs) could have resulted in a devastating cyberattack. Subsequently, section four describes low-cost mitigations which aim to mitigate the exhibited threats with various mitigation types described in each sub-section. Lastly, section five concludes the paper with closing thoughts.

2. The Oldsmar Water Treatment Attack

The attack began on February 5th, 2021, at 0800 where the attacker leveraged compromised credentials and the common remote access software TeamViewer to login to a plant operator's console. Following the login, the attacker immediately logged off and disconnected the session (Pinellas County, 2021). This was likely the attacker confirming system, connection, and credential validity. At this point the attacker could have also taken

a screenshot of the console supporting the planning of further actions. The console was manned by an operator at this time and the anomalous behaviour was mentally noted, but no action was taken. Later that day, at 1330 the attacker logged back into the operator console and used the plant's human machine interface (HMI) to command the increase of sodium hydroxide, also known as lye, to water which supplied thousands of people. Lye exposure can result in painful burns and permanent damage if ingested. The lye was increased from 100 parts per million (ppm) to 11100 ppm, an increase of over 100 times the intended amount. The operator observed this anomalous behaviour and commanded levels back to normal values. The operator then notified a plant supervisor, and the TeamViewer access was removed. This attack also occurred in the days leading up to the Super Bowl on February 7th which was being hosted 12 miles away from the facility, likely drawing more people to the area. According to county officials, it would have taken 24 to 36 hours of an increased mixture setting to have the malicious chemical mixture reach a point of distribution where it could have been consumed by customers. Additionally, county officials cited the use of pH level sensors with alarm reporting functionalities which would have notified operators of an imbalance (Pinellas County, 2021).

This attack lacks much of the complexity seen in previous cyberattacks targeting OT. Current reporting suggests the attacker did not attempt to evade detection and the attacker did not deploy custom scripts or programs, instead opting to leverage the existing operator interface to command their effect. While some would see this as an indicator of a low-skilled adversary, it is often the case that an attacker, regardless of skill, will use the easiest and fastest means to achieve a desired outcome. This can result in highly skilled attackers using techniques seen as simplistic or effortless. Therefore, it can be difficult to assess the attacker's skill in this case with the only potential indicator being the assumption that they did not meet their desired effect. Table 1 below breaks down relevant characteristics of the water treatment attack that occurred on 02/05/2021.

Table 1: The characteristics of the Oldsmar water treatment cyberattack

Attack Characteristic	***Oldsmar Water Treatment Attacker***
Attack Date	02/05/2021
Attack Time	1330
Access	TeamViewer Remote Service
Attack Execution	Process HMI
Detection Evasion	None Reported
Persistence	None Reported
Response Function Obstruction	None Reported
Operational Impact	▪ Momentary Control Manipulation ▪ Reduction of System Confidence

This table incorporates the current understanding of the attack's attributes and maps them to characteristics some of which are derived from the MITRE ATT&CK for industrial control systems (ICS) (Alexander et al, 2020).

3. Modification of Attacker Tactics Techniques and Procedures

Using the 02/05/2021 attack as a baseline and modifying certain attack characteristics generates two new theoretical attacker models. These two new attacker models vary by operational preferences for stealth and persistence but share the same mission of impacting the delivery of potable water, resulting in either an inability to deliver potable water or the delivery of dangerous water. Attack success is defined by the ability to command the increased addition of lye for the 24-to-36-hour timeline provided by county officials and the ability to degrade or eliminate the situational awareness provided by the system's pH sensors (Pinellas County, 2021). Given its proven success, all three attacker models will share a commonality of access via the TeamViewer remote service. A high-impact differentiator between the theoretical attacks and the 02/05/2021 attack is the date and time of attack execution. This temporal differentiator requires no additional skills to utilize and results in a delayed response and higher chance for mission success. The Oldsmar water treatment plant's normal hours of operation are publicly available via open-source information (City of Oldsmar: Water Division, 2021). Leveraging this information to launch attacks outside the normal hours of operation, both theoretical attackers will execute their operations just eleven hours and thirty minutes later compared to the 02/05/2021 attack, opting for an attack on 02/06/2021 at 0100 hours. Executing the attack at this time reduces the likelihood of an operator noticing the remote access and could increase the response time for lye level correction. For the sake of realism, neither theoretical attacker will leverage supply chain compromises, zero-day exploits, or low-level firmware manipulation as these concepts require a substantial amount of time and resources to employ. Table 2 below displays the various attacks and their TTPs.

Table 2: A comparison of attacker model characteristics.

Attack Characteristic	***Oldsmar Water Treatment Attacker***	***Theoretical Attacker A***	***Theoretical Attacker B***
Attack Date	02/05/2021	02/06/2021	02/06/2021
Attack Time	1330	0100	0100
Access	TeamViewer Remote Service	TeamViewer Remote Service	TeamViewer Remote Service
Attack Execution	Process HMI	Process HMI	Malicious Background Process
Detection Evasion	None Reported	Ransomware	System State Spoofing
Persistence	None Reported	None	Execution on Start-up
Response Function Obstruction	None Reported	Ransomware	▪ Command Blocking ▪ Alarm Suppression
Operational Impact	▪ Momentary Control Manipulation ▪ Reduction of System Confidence	▪ Control Manipulation ▪ Reduction of System Confidence ▪ Data Destruction ▪ Lengthy Inability to Deliver Potable water ▪ Possible Delivery of Unsafe Water	▪ Lengthy Control Manipulation ▪ Delivery of Unsafe Water

3.1 Theoretical Attacker A

Theoretical attacker A is impartial to attack discovery, instead opting to quickly impact the system while creating a difficult path for system recovery. Following the 0100 command to increase the lye amounts in the water supply, theoretical attacker A launches a ransomware attack on the system's computers to include operator stations. This ransomware attack reduces the operator's ability to control the system and can stifle the alerting of the pH imbalance due to a lack of situational awareness provided by their system interfaces. Communication could also be impeded if ransomware spreads to e-mail servers. A ransom note requesting bitcoin in exchange for decryption would manipulate responders into thinking the attack was financially motivated and concealing the attack's true intentions. Ransomware also has the added impact of encrypting log files, rendering them useless and increasing the difficulty of attack attribution.

According to a recent study, the number of reported ransomware attacks continue to increase yearly, one of which was the colonial pipeline ransomware attack (Reeder et al, 2021). This example illuminates the reality that malicious actors are targeting critical infrastructure systems which provide key services. This relevant example and continuing trend of ransomware use supports the realism of this theoretical attack scenario. Another emerging trend in ransomware is what is called "double-extortion" ransomware where attackers will apply additional pressure by threatening the release of sensitive stolen data in addition to the encryption of systems (Payne et al, 2021). To this end, it is possible that threats of malicious physical process manipulation carried out by dormant logic bombs will accompany future ransomware attacks on critical infrastructure to add additional pressure to the ransom. To illustrate the timelines associated with ransomware, a ransomware attack on the city of Baltimore took over a month to recover (Petcu, 2020). Acknowledging the differences in scale between the two systems, a report polling 2690 information technology (IT) professionals claimed that 66 percent did not believe their organization could recover from an unpaid ransomware attack in less than five days (Veritas Technology, 2020). Therefore, it is generally unlikely that full system functionality would be restored in the 24-to-36-hour timeline. Additionally, the odd hours of the attack delays initial attack discovery and subsequently the recovery process. Lastly, the unseen lye command and potential degradation of the system's pH sensing due to encrypted operator interfaces enables the potential delivery of unsafe water. In the event plant personnel discover the lye manipulation but are unable to access encrypted interfaces to assert positive control, a possible response could be to halt the delivery of water until systems are reconstituted. In both scenarios, the attacker has achieved their mission and impacted the quality of life for thousands.

3.2 Theoretical Attacker B

Theoretical attack B requires additional planning and domain understanding to execute but ensures guaranteed mission success. Attacker B relies on stealth to ensure that operators are unaware of the true state of their systems until dangerous water reaches customers. This attack requires a moderate understanding of cyber-physical systems and cyberattacks. This attacker gains initial system access via the TeamViewer remote access

application. Instead of continuing to use TeamViewer, attacker B establishes a malicious background process which provides subtle backdoor access unobservable to an operator under normal operating conditions. This process is configured to run at start-up, granting it persistence upon reboot cycles. This gives the attacker a long-term foothold to gain additional information on the system's processes, devices, and configuration. With this newfound system understanding, the attacker chooses to execute a man-in-the-middle (MITM) attack. This attack commands the system to add dangerous levels of lye to the water while reporting normal values to the operator. This can be accomplished in multiple ways and the attacker chooses whichever presents the highest chance of success based on initial reconnaissance.

In this theoretical scenario, the attacker chooses to modify the programmable logic controller's (PLC) ladder logic responsible for lye addition and pH reporting. PLCs are ruggedized real-time controllers prominent in many ICS environments which execute pre-programmed logic and interface with physical systems via control signals to actuators and sensors. The attacker reprograms the device, hard coding a dangerously high value of lye to be added which would otherwise be controlled by the operator or the device's original logic. The attacker also modifies the scaling of any reported values related to lye introduction and pH sensing to reflect normal values. This strips away awareness of the system's true status provided by the physical sensors connected to the PLC. An additional benefit that stealth provides is the inability to discern cyberattack from system failure. In the event plant personnel were able to detect the pH imbalance through an out-of-band channel, they would more likely attribute the anomaly to a faulty sensor or hardware failure then to a cyberattack. As a result, any hardware replacements would prove ineffective considering the attacker's foothold on the system. This attack uses long-term access and stealth to ensure the lye value's change lasts at least the 24 to 36 hours specified by county officials. It also considers the system's pH sensing and circumvents this via a MITM attack. Even if these operational requirements were not outright stated, the attacker's long-term foothold would have provided sufficient time to understand the system process and form attack requirements. Given this attacker's stealthy approach, it is highly likely that the first indications operators would receive of this attack would be calls associated with consumption of dangerous water, resulting in a successful mission for the attacker.

4. Theoretical Application of Potential Low-Cost Mitigations

The following subsections outline various solutions which attempt to mitigate the threats shown throughout this paper. These solutions prioritize threat mitigation while adhering to the budgetary and personnel constraints limiting many municipal utilities. Even so, all these solutions are directly dependent on the understanding of the current system and a lack of detailed system and configuration knowledge could reduce the impact of these mitigations. Additionally, any misconfiguration of these capabilities could result in a less effective mitigation solution. Lastly, these solutions are best utilized in combination with policies, procedures, and training that outline capability use.

4.1 Remote Access Limitations

The first mitigation is to establish and ensure a strong cyber-perimeter, this includes restricting and eliminating remote access wherever possible within operational parameters. Ensure devices are patched, prioritizing any internet facing devices. If remote access to plant data is required for situational awareness but positive control can be limited to on-site personnel, plants should consider implementing a one-way data-diode. If remote access is required for periodic remote maintenance, leave properly configured remote access systems offline, only bringing them online during coordinated maintenance time-windows to minimize internet exposure. Ensuring proper configuration and patching systems does not incur costs associated with new hardware, instead requiring minimal personnel time making it a cost-effective strategy to mitigate potential threats. One-way data diodes do incur hardware costs but are generally available for less than $10,000 and significantly reduce risk if they are the only operationally relevant solution for remote access removal at a given site. Proper remote access limitations would have mitigated the 02/05/2021 attacker and both theoretical attackers. Even so, while ensuring a strong cyber-perimeter is the first line of defence towards threat mitigation, it must be assumed that an attacker will still be able to gain access and that mitigation measures within the cyber-perimeter should be implemented. The following sections outline solutions which apply to internal systems and aim to mitigate threats within these boundaries.

4.2 Anomaly Detection and Prevention

A heightened awareness of the communications traversing the internal network can illuminate malicious actors on a network, aid in attack response, motive discovery, and attribution while assisting in the differentiation between hardware failure and cyberattack. While the 02/05/2021 attack was blatantly observed, the two

theoretical attacks are executed such that potential observation is minimized. In this case, properly configured network security monitoring and logging would have contained the artifacts and alerting necessary to begin responding to the attack. However, the effectiveness of this mitigation relies on how operationally relevant its detection mechanisms are as anomaly detection only designed for traditional IT system threats lack much of the domain-specific context present in OT systems. For example, an anomaly detection system comprised entirely of IT threat signatures would be unable to discern the lye increase as malicious. Therefore, a low-cost OT-relevant solution is required to maximize situational awareness in the event of a cyberattack.

Security Onion and ROCK NSM are two popular and free open-source solutions to network monitoring (Burks, 2012) (RockNSM, 2019). While these tools are IT-centric, additional rules can be added and customized to provide more OT-relevance. Digital Bond's Quickdraw is an example of free open-source rulesets which can be added to alert on OT threats (Peterson, 2009). Additionally, Malcolm is a free open-source solution which has additional support for the communication types prevalent within ICS environments (CISA, 2021). The most effective anomaly detection stems from rulesets generated by an understanding of process knowledge. For example, if the command to increase the lye amount should never exceed a certain value, a specific rule can be written to alert on this occurrence. These customized operational-specific rules require a deep understanding of the OT system and process paired with an understanding of the IT infrastructure required to implement them. This network monitoring solution requires minimal hardware investment, only a host machine to run the software and a configuration of network infrastructure to support port mirroring which provides data for ingestion. However, this solution does require a significant amount of personnel time to configure, install, and monitor. OT-tailored solutions which utilize artificial intelligence to automatically baseline a system and alert on deviations of standard behaviour exist, but they require a significant capital investment compared to the free open-source solutions but require far less personnel investment. Therefore, a solution choice should be determined based on a site's available resources. However, regardless of the solution, policies and procedures which outline actions following alert generation should be implemented to best utilize the capability.

In the event theoretical attack A is successful, and operators arrive to find machines presenting ransom notes requesting payment for system decryption, examination of the anomaly detection artifacts highlights indicators of when and how the attack occurred. If the anomaly detection has enough OT specificity, responders would be alerted of the lye change and the attacker's true motivations. If the alerting is configured to be sent to offsite personnel via e-mail or short message service (SMS) transmission the attack response becomes timelier, minimizing the impact of both theoretical attacks' off-hour execution. Lastly, due to the anomaly detection system's ingestion of mirrored traffic, it can be segmented from the network such that a ransomware attack would be unable to modify its artifacts.

Theoretical attacker B's stealth makes detection more difficult, but not impossible. Up-to-date and customized OT rulesets would detect the reprogramming of the PLC allowing plant personnel to attribute system abnormalities to a cyber incident. Additionally, the associated alerts would attribute the attack to a source device, allowing attack responders to segment and investigate the system with the malicious background process.

4.2.1 Bump-In-The-Wire Solutions

Anomaly detection and prevention can be backfitted into existing systems using bump-in-the-wire (BITW) solutions. These solutions are implemented using hardware which sits on the network in between devices. They monitor each communication packet traversing the network, alerting and blocking traffic which trigger pre-programmed security rulesets. Unlike the network monitoring solutions showcased in section 4.2 which used port mirroring to ingest data, the BITW solutions are positioned in the network to be able to act when a rule is triggered as opposed to just alerting an operator. These solutions require an investment in hardware to place throughout the system. The size of an installation will determine the significance of the hardware investment. However, if financial resources are limited, the installation of BITW solutions can be prioritized to critical systems.

In theoretical attack A, the command to increase the lye amount could be detected and automatically dropped. Accompanying alerts would give responders a timeline of events and provide insight into the attacker's true intentions. A similar ruleset prohibiting device reprogramming unless expressly allowed through operator coordination would eliminate theoretical attacker B's MITM attack, alerting reduces the adversary's stealth advantages, and provide responders with awareness to support further remediation.

4.3 Security Orchestration

Security orchestration automates many of the actions a network defender would take in combating a cyberattack via pre-defined playbooks. Orchestration could be a huge benefit to installations with limited personnel or no personnel dedicated to actively defending the network. Free orchestration solutions allow for a pre-defined limit of automated actions to occur within a time window, but these constraints would likely be within most requirements of a small municipal control system. Orchestration itself requires a minimal investment of capital and personnel resources, but it does require a system with security solutions implemented to be most effective. For example, orchestration could automate the generation and implementation of a new firewall ruleset based on observed malicious behaviour. This scenario would require a competent network monitoring solution to observe and alert on the malicious behaviour and would also require a firewall compatible with your orchestrator to ingest the new ruleset. If the orchestrator does not have sufficient data to inform its decision-making or the proper mechanisms to enact impactful defences its effectiveness is suppressed.

The ability to automate the response and adaptation to cyber threats in real-time is a valuable asset and can negate the benefits of both theoretical attacks' off-hour executions. Assuming a compatible security architecture is in place, the automation of the response process helps to ensure recovery occurs within the 24-to-36-hour timeline, greatly reducing the likelihood for attack success.

4.4 Ransomware Mitigation

With ransomware's increasing prevalence in recent years, solutions specifically tailored to minimize ransomware's impact should be considered. Offline backups assist with system reconstitution in the event of a ransomware attack. Due to the generally static nature of ICS environment configurations, storing offline backups of OT systems should be performed. As an example, the configurations associated with HMI displays and the logic files of the PLCs go unchanged for long periods of time. Offline backups of these systems are crucial in the event your systems are tampered with, or reprogrammed. The infrequent back-up cycle of OT components results in minimal personnel time investment and the likely small amount of data results in inexpensive offline storage requirements. The system's IT components should be backed-up more often if possible. Even so, ransomware recovery methods for the IT components which automatically backup systems on an unwritable opal drive would quickly speedup recovery time from weeks to hours (Challener et al, 2018). The capacity to greatly minimize the impact of theoretical attacker A's ransomware or theoretical attacker B's PLC reprogramming would allow operators to quickly reassert positive control and situational awareness. With positive control re-established, the lye levels would be discovered and corrected.

4.5 Training and Awareness

Training which attempts to increase the cyber-hygiene of personnel should be implemented to minimize unwitting insider threats. Periodic tests which help to understand the current state of cyber-hygiene within an organization should be regularly practiced. A common example of this being phishing e-mail tests. Additionally, personnel should be aware of the threat that cyber has on these systems and how they should respond given certain indicators or scenarios.

5. Conclusions

This paper described the 02/05/2021 attack on the Oldsmar water treatment plant and how it serves as an example for the current state of many systems responsible for providing basic needs. Theoretical scenarios showed how minor changes could have resulted in a dangerous situation for thousands of people. This paper's proposed mitigations increase the difficulty to execute these attacks while reducing their likelihood for success. The 2021-2022 Oldsmar annual budget contains only a single line directly relevant to this topic, "Continue network hardening measures around City's supervisory control and data acquisition at reverse osmosis water treatment plant and water reclamation facility." with no specifics describing how that will be accomplished or how success will be measured (City of Oldsmar, 2021). The shortcomings outlined in the paper are not isolated to Oldsmar and should be viewed as a general reflection of the municipal control system landscape. A landscape which could become a lucrative target for asymmetric cyber warfare operations. This paper hopes to be a motivator to augment existing systems with these mitigation concepts and serve as a potential influence for the requirements of the unbuilt systems of the future.

References

Alexander, O., Belisle, M. and Steele, J., 2020. *MITRE ATT&CK® for industrial control systems: Design and philosophy*, The MITRE Corporation: Bedford, MA, USA.

Burks, D., 2012. *Security onion*.
Challener, D.C., Kruus, P.S., Fink, R.A. and Farlow, J.F., Johns Hopkins University, 2018. *Apparatus and method for preventing access by malware to locally backed up data,* U.S. Patent 10,049,215.
Cybersecurity and Infrastructure Security Agency (CISA), 2021, *Malcolm*, [online], https://github.com/cisagov/Malcolm
City of Oldsmar, 2021. *City of Oldsmar Fiscal Year 2021/2022 Annual Budget*, p 111.
City of Oldsmar: Water Division, 2021. *Hours,* Oldsmar, FL, USA.
Department of Homeland Security, 2015. *Water and Wastewater Systems Sector-Specific Plan,* District of Columbia, USA.
Heenan, R. and Moradpoor, N., 2016. *Introduction to security onion.* The First Post Graduate Cyber Security Symposium.
Payne, B. and Mienie, E., 2021. *Multiple-Extortion Ransomware: The Case for Active Cyber Threat Intelligence.* In ECCWS 2021 20th European Conference on Cyber Warfare and Security. Academic Conferences Inter Ltd. p 331.
Petcu A. G., 2020. *The Curious Case of the Baltimore Ransomware Attack: What You Need to Know*.
Peterson, D., 2009. "Quickdraw: Generating security log events for legacy SCADA and control system devices", *Cybersecurity Applications & Technology Conference for Homeland Security* (pp. 227-229). IEEE.
Pinellas County Sheriff's Office, 2021. *Treatment Plant Intrusion Press Conference,* Oldsmar, FL, USA
Reeder, J.R. and Hall, C.T., 2021. *Cybersecurity's Pearl Harbor Moment: Lessons Learned from the Colonial Pipeline Ransomware Attack.*
RockNSM, 2019. *What is ROCK*.
ThreatLocker Incorporated, 2021. *Protecting Water Infrastructure Against Cyberattacks,* Maitland, FL, USA.
Veritas Technologies LLC, 2020. *The 2020 Ransomware Resiliency Report,* Santa Clara, CA, USA.