This article has been accepted for publication in IEEE Security & Privacy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSEC.2023.3264595

Automating recovery in mixed OT/IT critical infrastructures

Sebastian Obermeier
Lucerne University of Applied Sciences and Arts

Thomas Jösler
Lucerne University of Applied Sciences and Arts

Stephan Renggli
Lucerne University of Applied Sciences and Arts

Maurus Unternährer
Die Mobiliar - Versicherungen und Vorsorge

Bernhard Hämmerli
Lucerne University of Applied Sciences and Arts

***Abstract*—Cyber incident recovery in modern power systems is challenging because OT and IT components are merged in one network. Efficient recovery demands high degrees of automation with as little as possible recovery time.**
To tackle automated recovery, we have utilized our lab that resembles a substation automation system including contemporary security controls and an integration of OT/IT components. Whenever this lab is used to train cyber security professionals, the lab state changes, especially after cyber-attack exercises. Thus, the lab needs to be recovered into known good states, comparable to real systems after a (potential) cyber-attack.
In this contribution, we present various concepts for backup, recovery, and configuration management. We evaluate these concepts and identify limits for applying them to OT systems. We observe that automated recovery drastically speeds up recovery. However, while most IT components are well prepared for automation, OT hard- and software is not and imposes hurdles like forced interactive installations or license activation due to the legacy character.
Finally, we discuss that automated recovery can provide benefits for real-world environments but needs further research, and present a cost/benefit analysis for the laboratory.

Index Terms: cyber security substation automation lab automated recovery

1. Introduction

"The National Risk Analysis of Disasters and Emergencies in Switzerland" has identified that the highest risk for the population is a long energy shortage [11]. For strengthening the national infrastructure, cyber security education and research also focuses on energy systems, in particular substation automation systems. To educate not only on a theoretical level but also through delivering practical experiences, we have built a cyber security substation automation lab that follows a real-world substation. The lab is used for education of students and professional engineers operating critical infrastructures.

The main idea of this lab is to provide a hands-on experience on how a substation can be protected following various standards and best practices, and also serve as a base for further

research. For this purpose, the lab implements a modularized approach for cyber security controls within critical infrastructure environments. The lab starts with a substation scenario that employs no cyber security controls at all. The students run hacking-tools, perform inventorying, and interact with the substation on both, network and physical level. They determine the initial maturity of the environment through a security assessment and identify by using standards and best practices which cyber security controls should be introduced as next step. After each iteration, they verify the effectiveness through security tests and re-assess the system. This approach allows to acquire various important competencies.

1.1. Problem Statement

The lab will undergo several "cyber-attacks" per week during training sessions (including malware attacks) and needs to be recovered to a set of defined states. As the training participants can damage the system (and even must do in some exercises), the system must be recovered completely and built from scratch after each training session.

Another challenge is the timing of the lab offering, which is targeted to be offered on five consecutive days, requiring to recover the lab infrastructure after each training day. As this is a complex task and time-consuming if done manually, the recovery processes for the whole lab must be automated to the greatest extent possible to be able to offer such a lab setup.

However, there is no guidance for such automation: In a non-lab environment, recovery is seldom, done manually, and targeting a single known-good state only. Therefore, system integrators will not provide any features on automating this task such as pre-built scripts or configured virtual machine (VM) images. Even if VM images would be supplied, certain components are time-sensitive (e.g., Active Directory domain controllers). If the time on two domain controllers differs to much, they will no longer be able to synchronize. The same applies to update sequence number (USN) rollbacks caused by restoring an older Active Directory state than that is present on other domain controllers. This affects the feasibility of possible recovery methods. According to Microsoft's documentation, it is not advised to use virtual machine snapshots for domain controllers because of possible synchronization issues[1]. If disk images are used for recovery, the backup & restore software used must be able to handle USN rollbacks.

Manufacturers of devices follow cyber security standards, which demand incident management and business continuity, but do not require automated recovery strategies. Therefore, devices are not optimized for automated recovery and impose hurdles like software activation or forced interactive installations that hinder an efficient automated recovery process. Many of the existing labs are mostly used for research and have not tackled this problem either.

The concept of configuration management [2] includes the management of various system configurations.Although some ideas and technical solutions can be used for our purposes, configuration management does not focus on restoring complete systems that are compromised by malware. In our case, however, some exercises involve students to apply offensive security tools, resulting in unpredictable states of the system that requires a complete restoration.

1.1.1. Contributions

We present a lab concept for substation automation systems and give guidance on the automated restoration of the lab into different states. Thus, the contributions of the paper are:

- It shows how the recovery process can be automated by presenting an arsenal of backup, recovery, and configuration management tools and concepts.
- It paves the way for complex training scenarios and exercises that completely change a lab's infrastructure and security controls, which would not been economically and timely possible without an automated recovery into different states.
- It discusses how automated recovery can be leveraged in real-world environments and identify obstacles that hinder an efficient restora-

[1]Microsoft: Virtualizing Domain Controllers using Hyper-V (2022), https://docs.microsoft.com/en-us/windows-server/identity/ad-ds/get-started/virtual-dc/virtualized-domain-controllers-hyper-v

[2]ANSI/EIA-649-C: Configuration Management. Stan- dard, SAE International, Pennsylvania, USA (2019)

This article has been accepted for publication in IEEE Security & Privacy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSEC.2023.3264595

tion during emergency situations, for which preparation should be conducted before an incident happens.

2. Related work

Several lab approaches for education and research that describe how a lab can be setup and operated exist, e.g., [4], [1]. However, they do not focus on recovery strategies and how to rebuild the system into different states of the lab. Building blocks to solve the problem described can be found in other areas, e.g., those that deal with cyber attacks.

In incident response, several strategies to automatically tackle cyber attacks exist, e.g. [5] presents a heuristics based approach to automatically determine an efficient set of countermeasures for an attack. [12] focuses on dynamic snapshots in cloud environments. [15], [3] present automated decision making approaches for incident management. Using such approaches in a lab environment can help to automatically investigate the incident and the changes caused by the training participant, which can be helpful to determine an optimal set of resources that needs to be recovered. However, in the worst case, the complete system has to be reset anyway.

A solution for automated backups of communication networks is presented in [6], but not for restoration. [10] discusses utilities for backing up individual software and provides instructions for restoration. In industrial systems, frequent automated backups for device configuration may not be necessary due to system stability.

An automated recovery through virtual machine restoration after an "Internet worm" has attacked a system is presented in [8]. The approach ignores all data packets after the restoration and thus helps in preventing zero day attacks.

With respect to power systems, CISA has created a national guideline for surviving a power outage, stressing not only the the importance of recovery, but also the need for efficient decision making processes in case of cyber attacks [9]. Recent alerts indicate that "traditional approaches to securing OT/ICS do not adequately address current threats", as national groups and Advanced Persistent Threats (APTs) are targeting critical infrastructure [2]. As especially targeted APTs can anchor themselves deep in the system, a full recovery into a known good state is essential for real-world installations. [14] suggests an approach for determining the optimal re-closing times in case a cyber attack has caused a disruption of the power transmission. However, this paper does not focus on the recovery of individual devices and systems, especially in case malware has compromised them.

Software defined networking (SDN) is used in [13] to create a flexible power system security lab. Logical components, such as the SDN controller or the human machine interface (HMI), are being run on a ESXi hypervisor. While this lab is not focused on backup, our approach could complement such solution.

Ansible[3] is an open-source tool for software provisioning, configuration management, and application-deployment, which enables infrastructure as code (IaC). In [7], an approach to manage the computer labs at Brno University of Technology is demonstrated, which includes the setup of different operating systems using Ansible. However, the authors do not disclose any backup or recovery considerations.

To summarize, we have found some building blocks that can help in automating the recovery of a critical infrastructure lab, but no detailed description solving the problem.

3. Lab Structure

The substation automation lab, illustrated in Figure 1, is set up on the basis of 220kV high-voltage substations (called "DOMPROD") and a central control station (called "DOMOT"). Currently, it consists of six substations, each installed in one rack, and a central control station. The setup is designed for portability and scalability: All racks are portable and the number of substations can be extended through replicating the racks.

3.1. Lab Components and Configuration

The lab utilizes a mix of hard- and software components. Each of the racks contains an HP DL20 Gen10 server which serves as ESXi host and executes different virtual machines. The network control center functionality is implemented in a single rack and connected to the individual

[3] https://github.com/ansible/ansible

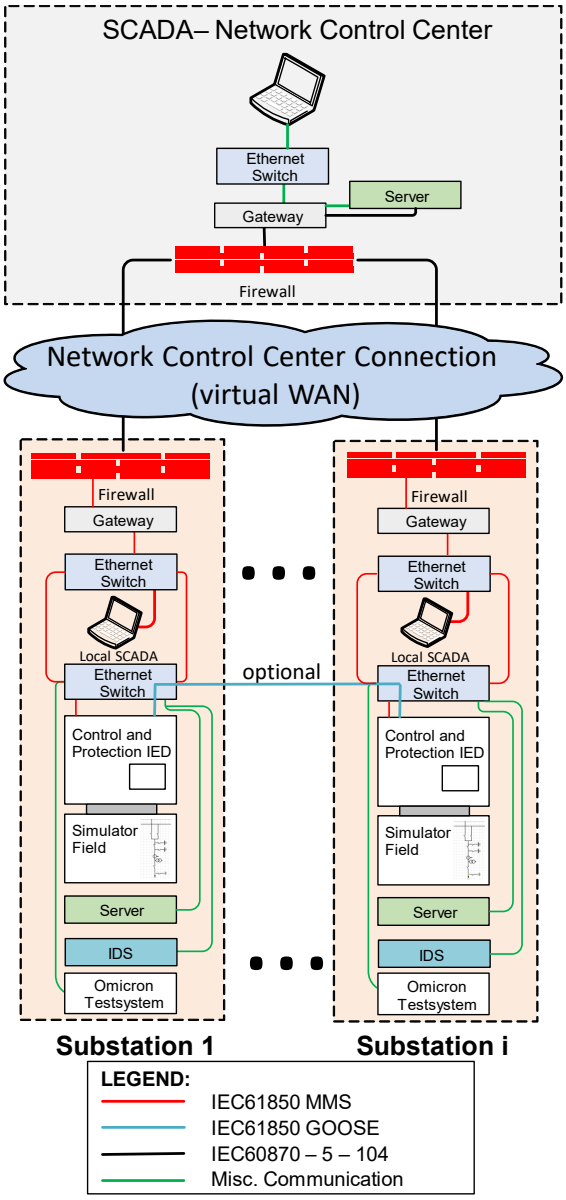


Figure 1. Lab network setup and protocols

substations through the IEC60870-5-104 protocol.

Within each substation, the IEC61850 Ed. 2 MMS communication protocol is used, e.g., to perform switching operations. Optionally, a modified setup can use IEC 61850 GOOSE communication through linking two or more IEDs, as indicated by the blue line in Figure 1. However, for this setup, the switch needs to support GOOSE communication, which is not always the case for IT products. The Omicron Testsystem (CMC 356) in the bottom box is used to verify the protection functionality of the system. It can create, for example, characteristics that occur when a lightning strikes a power line.

The components of the lab are listed in the taxonomy of Hardware and Software in Table 1. We employ typical components that are also used in real-world installations, except the use of virtual machines, which is not common in substations, but we argue in Section 6.2 that there are benefits for real-world installations as well.

To manage the devices, we use the official vendor configuration software, i.e., ABB PCM600 and ITT600 for the IED and Streamconsole for the Elvexys XPG gateway.

3.2. Special OT Challenges for Automated Recovery

We have encountered various challenges for automated recovery while building the OT security lab.

OT software like the used SCADA system COPA-DATA Zenon or ABB SDM600 require special licensing. Zenon uses software licensing, and each license has to be registered online with Copadata. If the license is no longer used, it has to be deactivated online. If a hardware failure or other circumstances that prevent access to the virtual machine that hosts Zenon occurs, the license is lost and needs to be reacquired with the help of COPA-DATA support.

COPA-DATA Zenon supports silent installations and even accounts for installation in virtual machines using special checks during the installer routine. The version of the Zenon installer used in our lab includes a bug which prevents the software from being installed in unattended mode via WinRM. As workaround, Zenon is installed via a PowerShell script which has to be executed manually in a regular user session (e.g. Remote Desktop).

ABBs SDM600 uses a combination of software and hardware licensing. Each installation requires a specific USB dongle connected to the virtual machine that has to match the software licensing file that needs to be present in the programs installation directory.

Elvexys XPG Gateway and ABB RED670 use proprietary software for system configuration (StreamConsole and PCM600, respectively). Nei-

Table 1. Taxonomy of Software and Devices

Rack	Category	SW	HW	Component	Description
Control Center	SCADA	✓		COPA-DATA Zenon	SCADA system of the network control center
Control Center	Server		✓	ESXi Host	HP DL20 Gen10 server hardware
Control Center	Data Management	✓		SDM600	Hitachi Energy System Data Manager
Control Center	Directory Server	✓		Active Directory	Microsoft User Account Management
Control Center	Recovery	✓		Ansible	Ubuntu Linux Ansible Server
Control Center	Configuration	✓		ABB ITT600	Toolbox for Substation Tests
Substation	SCADA	✓		COPA-DATA Zenon	Local substation SCADA system
Substation	Server		✓	ESXi Host	HP DL20 Gen10 server hardware
Substation	Protection Relay		✓	ABB IED 670	Protection relay
Substation	RTU		✓	Elvexys XPG	Protocol converter and gateway for communication with the control center using IEC104
Substation	RTU	✓		Streamconsole	Configuration software for Elvexys XPG
Substation	Data Management	✓		SDM600	Hitachi Energy System Data Manager
Substation	Configuration	✓		ABB PCM600	IED configuration
Substation	Fault Analysis	✓		Comtrade viewer	Siemens viewer to visualize power grid disturbance records
Substation	Test Instrument		✓	Omicron CMC 356	Relay Testing for power amplification
Substation	Test Instrument	✓		Omicron Test Universe	Test cases for CMC 356
Substation	Remote Desktop	✓		Royal TS	Client for remote desktop sessions
Substation	Network analysis	✓		Wireshark	Tool to analyze network traffic
Substation	Data transfer	✓		WinSCP	SFTP- and FTP client software
Substation	Test Instrument		✓	REC670 Simulator	ABB Switchbox to simulate field equipment
Substation	Security		✓	StationGuard	Omicron Intrusion Detection System

ther software includes a command line interface which could be called via IaC or configuration script.

Software, which uses the Windows Security Identifier (SID) of the NETWORK and SYSTEM accounts for authentication purposes, such as Microsoft SQL Server, cannot be easily cloned because the cloning procedure (sysprep) often resets SIDs, which then breaks the internal authentication of the mentioned software products. Zenon, SDM600 and PCM600 depend on Microsoft SQL Server. While Microsoft did add sysprep support for empty Microsoft SQL server instances in Microsoft SQL Server 2012, it is still unsupported for already configured systems[4].

Infrastructure as Code or scripting in general requires software installations without manual interaction. Usually, this is achieved by calling a setup executable using a silent parameter. Whereas IT software typically provides silent installation options by default, some OT software does not. Examples are ABBs SDM600, Siemens COMTRADE Viewer and Omicrons Test Universe. Siemens COMTRADE Viewer and Omicrons Test Universe both use a setup .exe as a wrapper for various MSI installers contained within the .exe file. By observing the MSI installation parameters used by the exe-wrapper through log file analysis, it has been possible to recreate the installation routine using silent installation parameters when calling the MSI files. ABBs SDM600 cannot be installed silently, which has been confirmed with ABB support. The mentioned reason are parameters that have to be entered in the setup GUI. It is not possible to feed these parameters into the setup using the command line.

The PCM600 IED device configuration software does offer silent installation, however, due to timing issues within the installer, this seems to be unreliable. Another difficulty concerns the PCM600 project files. Project files can not be created/configured using IaC templating due to the proprietary binary file format used. Therefore, for each substation, a separate project file needs to be maintained. The configuration of the IED using this project file can also not be automated, which is still a manual task.

4. Backup and Recovery Concepts for Labs

In the following section, we describe the general backup and recovery concepts as well as tools that serve as base for implementation in our lab environment. In Section 5, we describe our recovery strategy while in Section 6, we evaluate the applicability of these approaches in our lab

[4]Microsoft: Install SQL Server with SysPrep (2022) - SQL Server, https://docs.microsoft.com/en-us/sql/database-engine/install-windows/considerations-for-installing-sql-server-using-sysprep

environment.

Table 2. Components and their reset requirements

Component	Reset requirement
ESXi hosts	no
Virtual Machines	yes
vSphere Appliance	no
Firewall	yes
Switch	yes
Microsoft Active Directory	yes
SDM600	yes
PCM600	yes
Zenon	yes
XPG Gateway	yes
WinSCP	yes
Royal TS	yes
Windows Server 2019	yes
NTP Server	no
Ansible Management Server	no
DLCache	no
ABB RED670	yes

In general, all components which students are able to interact with need to be reset between lab sessions. While not all components are used in every lab exercise, it is still possible that curious students change settings or delete files on components that are not part of the ongoing exercise. To allow a consistent experience for each of the student groups and to guarantee a working lab environment, all components that students interact with must be reset between lab sessions. Table 2 illustrates which components need to be reset.

4.1. Disk Images

Disk images are files containing the structure and data of a complete disk on a block level. Creating and restoring disk images requires specialized software and can not be done on a running system. Examples for such software are Acronis True Image or Veeam. Disk images are immutable once created. This has advantages in terms of security such as the ability to hash the file to verify its integrity.

In a lab environment however, this is not required as frequent modifications of the image to adapt the lab for new exercises is not common. Another disadvantage of disk images is the need for external documentation as the image itself cannot be examined (in terms of included software and configuration settings) by users. Cloning disk images comes with restrictions because of its one-to-one copying nature. Some restrictions that need to be solved by the backup & restore software are hostname changes, IP addressing and User SIDs which are hardcoded in some software installations (e.g. Microsoft SQL Server).

4.2. Configuration Files

Devices such as firewalls and switches usually load their configuration from an internally stored configuration file. The configuration file format varies from platform to platform. Some configuration files use proprietary configuration files while others implement well-established standards such as JSON or XML. Once all configuration has been made the file is exported. To restore a specific device configuration state the configuration needs to be imported into the device. Common ways to import configuration files are Web GUIs, API Calls, Command-Line-Interface or USB devices. Configuration files can be easily changed if requirements change in the future. Configuration files can also be seen as a way of documenting all necessary settings. The feasibility of configuration files as documentations depends heavily on the readability of the file for humans.

4.3. Virtual Machine Snapshots

The concept of virtual machine snapshots only applies to virtualized systems. After installations and configurations, a snapshot of the virtual machine state including memory is taken. Restoring a snapshot overwrites memory and discards disk changes. Snapshots cannot be altered once created. Non-virtualized systems may have interoperability issues if interacting with restored machines due to out-of-date memory snapshots, which may lead to sequence number problems.

4.4. Automated Docker creation and installation

Using Docker, pre-configured docker containers can be created. This approach works well, if a large number of identical instances are required. While Linux based containers have achieved feature-parity compared to bare metal installations or virtual machines, Windows containers still have some limitations, especially when using GUI applications. Docker provides easy to use configuration files (Dockerfile, docker-compose.yml) which can also be used for rudimentary documentation purposes. Docker configuration files are also easily extendable.

This article has been accepted for publication in IEEE Security & Privacy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSEC.2023.3264595

4.5. Script-based Configuration

Based on a script, a certain configuration is set on a target device or machine. Script-based configuration is common on operating systems such as Linux (bash scripts) and Windows (PowerShell scripts). Scripts run their configuration commands without assessing if the targeted system is already in the desired state. Depending on the case this can lead to unwanted system states. One example are duplicate database entries when running the same database restore script multiple times. Scripts provide an easy to use method of configuring a given system while also implicitly documenting all changes. Scripts can be easily modified and extended at a later stage if need be.

While some network devices allow for internal script execution, this is not possible on most devices. Most OT devices do not support script execution.

4.6. Infrastructure as Code

Using Infrastructure as Code (IaC) the desired configuration of all components of a given environment such as physical hosts, network devices, virtual machines and software are represented as code. The IaC runtime (e.g. Ansible) then performs the necessary configuration changes to achieve the desired configuration. Changes are performed only if they are required, this ensures idempotency. IaC allows for centralized modifications which are then applied to all affected devices. IaC Code is written in a human-readable format which at the same time documents all changes that will be implemented on the target systems.

IaC requires all target components to be configurable remotely. Usually this is achieved using SSH or WinRM, depending on the target OS. For network devices CLI interaction is used. Interfacing with virtualization platforms such as VMware vSphere is achieved via API communication using HTTP/S.

IaC allows for a large number of hosts to be configured simultaneously, this speeds up deployment considerably.

4.7. Applicability

The matrix in Table 3 displays the applicability of different recovery methods for components which need to be reset between different lab sessions.

5. Lab Recovery Implementation

Currently, the IaC tool Ansible is used for lab configuration and recovery. A virtual Ubuntu server has been set up in the DOMOT rack, i.e., the network control center, which serves as internal Ansible server. The required Ansible code repository is cloned to the server and an Ansible setup role is run. The Ansible setup role downloads all required files, installers and dependencies to the local Ansible server.

First, all necessary VMs need to be deployed on all substation ESXi servers. After the deployment is finished, a pre-configuration VM snapshot is taken. For subsequent lab resets the general sequence looks as follows:

1) Reset VMs to pre-configuration snapshot for DOMOT and DOMPROD
2) Reset Firewalls and Switches in DOMOT and DOMPROD
3) Configure DOMOT Active Directory Domain Controller (Parent)
4) Configure DOMPROD Active Directory Domain Controllers (Child)
5) Configure other DOMOT and DOMPROD VMs.

We have created a central Ansible repository that includes all playbooks for the tasks mentioned above. For each function (e.g., AD, SCADA, Remote Administration), a dedicated playbook has been created for each required state. Each playbook contains roles which in turn contain tasks that are executed on the specified hosts.

The playbook shown in Listing 1 is our central playbook containing all required roles to set up the final state for the Remote Administration hosts. It includes the full set of the lab's cyber security features. Some roles are only applied on substation or control center remote administration hosts.

The playbook is executed sequentially from top to bottom. Some roles are called while specifying additional variables (vars:, e.g. line 9). These roles are reusable and can be used for DOMOT and DOMPROD hosts, reducing code repetition. Other roles such as windows_activate

This article has been accepted for publication in IEEE Security & Privacy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSEC.2023.3264595

Table 3. Applicability matrix for different restore concepts to lab components

	Disk Images	Config. Files	VM Snapshots	Docker	Config. Scripts	IaC
Virtual Machines	✓	×	✓	×	✓	✓
Firewall	×	✓	×	×	✓	✓
Switch	×	✓	×	×	✓	✓
Active Directory	✓[1]	×	×	×	✓	✓
SDM600	×	✓	✓	×	×	×
PCM600	×	✓	✓	×	✓	✓
Zenon	×	✓	✓	✓[2]	✓	✓
XPG Gateway	×	✓	×	×	×	×
WinSCP	✓	✓	✓	×	✓	✓
Royal TS	✓	✓	✓	×	✓	✓
Windows Server 2019	✓	×	✓	×	✓	✓
ABB RED670	×	✓	×	×	×	×

[1] software needs to be USN rollback aware [2] only Zenon runtime

```yaml
# this playbook sets up all sotslra vms
---
- name: 'setup all sotslra vms'
  hosts: 'windows_remoteadmin'
  gather_facts: true

  roles:
    - role: 'dns_server_set'
      vars:
        dns_server_set_ips: '{{ dns_server_ip }}'
    - 'windows_activate'
    - role: 'ad_domain_join'
      vars:
        ad_domain_join_domainname: '{{ domain_name }}'
        ad_domain_join_domainadmin_user_username:
            '{{ domainadmin_user_username }}'
        ad_domain_join_domainadmin_user_password:
            '{{ domainadmin_user_password }}'
    - 'royal_ts_install'
    - 'royal_ts_configure'
    - 'wireshark_install'
    - 'rsat_install'
    - 'ied_explorer_install'
    - 'stream_console_install'
    - 'omicron_testuniverse_install'
    - 'siemens_comtradeviewer_install'

- name: 'sotslra domot specific configurations'
  hosts: 'windows_remoteadmin_domot'

  roles:
    - 'winscp_install'
    - role: 'winscp_scadagw_configure'
      vars:
        winscp_scadagw_fqdn:
        "{{ hostvars['sotslgw.domot.lab'].inventory_hostname }}"
        winscp_scadagw_username: 'labadmin'

- name: 'sotslra domprod specific configurations'
  hosts: 'windows_remoteadmin_domprod'

  roles:
    - 'helinks_sts_install'
    - 'abb_pcm600_install'
    - 'abb_pcm600_configure'

- name: 'activate rdp on all sotslra hosts'
  hosts: 'windows_remoteadmin'
  gather_facts: true

  roles:
    - role: 'rdp_enable'
      vars:
        rdp_enable_groupnames_add:
          - '{{ domain_name }}\Domain Users'
```

Listing 1: Ansible Playbook for Remote Administration Host

This article has been accepted for publication in IEEE Security & Privacy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSEC.2023.3264595

(line 11) do not require additional parameters as they are generic by nature and work on any (Windows) host.

Another major factor that needs to be considered when designing playbooks is time optimization. The following changes were implemented to optimize the playbook execution duration:

- reset to VM snapshots instead of recreating VMs
- organize Ansible playbooks by function instead of location
- implement download caches to minimize necessary file transfers via network
- configuration dependencies and tasks that slow down execution in general (e.g. reboots)

6. Evaluation

We have implemented the recovery concept in our lab environment and evaluated it through use in multiple runs of the different lab lectures, resulting in the need to reset the lab on short notice between lab sessions.

The lab environment can be setup or restored to a working state in about four hours time. With the exception of four components, all recovery steps could be automated and run in an unattended fashion. The remaining manual restoration steps require one hour of manual configuration time per rack.

During the initial setup of the lab, in which the recovery system was not available, we had to perform a manual setup and configuration, resulting in an overall effort for all racks of around 56 hours (14 times as long). The use of IaC also mitigates human errors in the process, resulting in all racks being configured consistently each time the reset is performed.

Table 4 shows an overview of the automation possibilities. While all IT components can be reset smoothly, this does not apply to four of the OT components, namely ABB SDM600, RED670, XPG Gateway, and Zenon.

6.1. Limitations

For four components, we were not able to perform an automated recovery due to the reasons below. The solution is to either not reset these components or to perform a manual restoration.

XPG Gateway and RED670 cannot be configured and/or restored using IaC. Backup and restore functionality is only available when using the proprietary GUI-only software. For the time being XPG Gateway and RED670 need to be setup and restored manually.

Table 4. Evaluation of the configuration automatability of different components and software

Component	IT / OT	Automation
Virtual Machines	IT	supported
Firewall	IT	supported
Switch	IT	supported
Active Directory	IT	supported
SDM600	OT	not supported
PCM600	OT	(supported)
Zenon	OT	(supported)
XPG Gateway	OT	not supported
WinSCP	IT	supported
Royal TS	IT	supported
Windows Server 2019	IT	supported
RED670	OT	not supported

ABB SDM600 does not support silent installations. Additionally, when installing the software manually there seems to be a timing issue between two parts of the installer (dependency during installation handled by waiting for 20 seconds) which sometimes causes the installation to fail. Automated configuration restore using IaC is not possible. These features are only available in the GUI of the software itself, no CLI is offered. For the time being SDM600 has to be installed and restored manually.

For the following two components we were able to create workarounds to partially automate the recovery process. While this does work for the moment there is still potential for optimization.

As described in Section 3.1, Zenon needs to be installed using a PowerShell script which is executed manually in a regular user session.

The solution to the PCM600 install reliability problem is to simply rerun the installer until the installation succeeds. While this solution allows for unattended recovery, it is not ideal.

6.2. Applicability to real-world installations

The lab has been constructed in such a way that is resembles a state-of-the-art substation with devices, software, and protocols that is used throughout the industry. It has been installed and configured by a dedicated engineering company that also works for national grid operators and has used industry standards for configuration of the system. The amount of simulation is limited to the power characteristics and primary physical

equipment.

Therefore, the setup and reset approach shown in this work is not limited to lab environments but can also be theoretically applied in productive systems. However, up to now, the authors have not seen such degree of automated (re-)installation, virtualization and automation in real live substation installations as in the laboratory. Before using the approach in a real productive environment, additional research and evaluation is needed, especially for safety reasons. To evaluate the benefit, the following attack model is used to illustrate how the approach of automated recovery mitigates attack consequences by speeding up the recovery process.

6.2.1. Attack Model The following attacks on a substation are considered:

- A1: A ransomware has infected the station computer and encrypted all files.
- A2: An Advanced Persistent Threat has infected the engineering PC or station computer.
- A3: An attacker has physically entered the substation and has made malicious configuration changes.
- A4: A malicious firmware has been installed on the IEDs or other embedded systems.
- A5: An attacker has installed a rogue devices in the network and performs network attacks.

6.2.2. Applicability

The following table shows for which kind of attacks the method is applicable, i.e., where is speeds up the process and assists in recovering the system into a known good state. We assume the system for recovery is shut down during regular operations.

Attack	Applicability of the approach
A1	✓, assuming the recovery system is still operational
A2	✓, except for BIOS/UEFI attacks
A3	(✓), for auto-configuration
A4	✗, no automatic firmware restoration
A5	✗, not possible

6.2.3. Hurdles There are hurdles to apply the concept in real-world installation. One of the challenges is that components must allow automating installations and configurations (or at least not prevent it), which is not common in OT environments. Especially software-based license key activation that requires a connection to the software manufacturer can become a major obstacle in an incident case when the system must be restored to an operational state as fast as possible. Some of the real-world software we have used requires such license activation.

Another reason why the concept is not widely used is that OT environments are focused on availability. Executing multiple VMs on a single hardware server reduces hardware redundancy. Although this can be overcome by using a second hardware appliance and fail over protocols, practical experience is missing. A benefit is that our work does not only apply to cyber security incidents, but can also improve restoration time in case of hardware failure.

Other benefits include almost immediate recovery from attacks and a known good state (exceptions are BIOS or hardware-destroying attacks). However, OT software is not ready for full automation just yet. First, installers need to work more reliably and provide silent installation switches. Second, installed software needs to offer command-line interaction possibilities for operations such as configuration backups and restores and license activation. Another necessary improvement would be human-readable configuration files. Standards like XML and JSON allow for the use of templating which results in only having to maintain one configuration file instead of maintaining a separate configuration file for each rack.

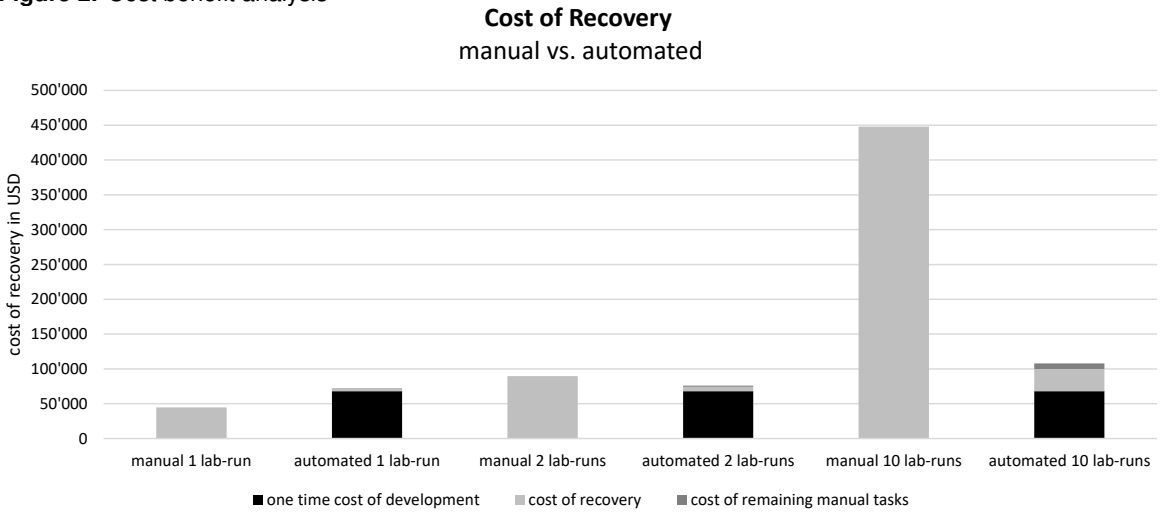
Although the approach shown in this work has many benefits, the development costs are not the be neglected. In the next section, we show a cost benefit analysis, which helps in determining the effort and to decide on the concept.

6.3. Cost-Benefit Analysis

The cost-benefit depends on the factors time and number of repetitions. The factor time involves the reduction in man-hours and the reduction in overall system downtime. In addition, the number of instances, the number of recovery cases and their probability of occurrence have an influence. As multipliers in the formula, these

This article has been accepted for publication in IEEE Security & Privacy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSEC.2023.3264595

Figure 2. Cost benefit analysis



have a leverage effect on the result.

In the lab environment, the probability of occurrence of a total disaster recovery is 100%. Likewise, the number of repetitions per lab-run is specified. In a real-world scenario, these parameters are difficult to quantify. Simplified, a proactive development of an automated recovery strategy is worthwhile in two cases. Firstly, if the probability of occurrence is high and secondly, if the cost of downtime is high.

For the lab environment, the calculation results in cost savings of USD 40'800 per lab-run, which consists of five consecutive training exercises on seven racks. In total, 35 recovery steps are required during one lab-run.

The following equation describes the cost-benefit analysis.

$$\text{Benefit} = \Big((\, C_m - C_a) \times i + (D_m - D_a) \times L_h \Big) \times R \times P - F \tag{1}$$

where:

C_m = cost of manual recovery
C_a = cost of automated recovery
i = number of instances
D_m = duration of full manual recovery in h
D_a = duration of full automated recovery in h
L_h = monetary loss per hour of system failure
R = number of repetitions
P = probability of occurrence
F = fixed cost of development

A factor that is not priced into the formula provided is the overall lab utilization. A lab that can be offered on five consecutive days yields higher revenues due to better utilization of resources than a lab that cannot be offered every day due to manual recovery downtime.

Figure 2 illustrates the cost benefit analysis for our lab. The break-even is hit after 1.7 lab-runs. If the lab will be expanded in the future by adding additional substations, the cost savings will increase linearly.

7. Summary and Conclusion

To educate on a practical level about cyber security in energy systems, a hands-on lab resembling a power substation has been created. To provide participants a working lab state for each lab session, a solid recovery mechanism is required.

In this article, different recovery concepts for a typical power substation and its components have been examined. Infrastructure as Code has been evaluated as the most effective method to restore the lab infrastructure. While all our lab's IT components can be restored automatically, some OT software and OT hardware need manual steps for a full restoration. The main hurdles for automated recovery of OT software we have identified concern license activation, automated installations, and unreliable configurations. For some components, workarounds for semi-automated recovery have been implemented, for others, manual steps during recovery are still

necessary.

Our evaluation has shown that the effort for an automated recovery was paid back in less than two lab runs and allows for a dense utilization of the lab. In a non-lab scenario, additional societal costs for downtime must be considered as well. We argue that the proposed concept can provide benefits for commercial energy suppliers with respect to quick and reliable recovery, but must be evaluated before used in real-word systems.

Future work includes optimization of automation processes and speeding up the lab recovery procedure. While some limitations such as interactive installations cannot be overcome until a new software release is published, investigation into other aspects such as unreliable installations will be conducted. The authors also plan to verify the developed recovery concept in a real-world substation to evaluate the advantages of the concept when leveraged in today's OT industrial processes.

Finally, we believe that professionals in the OT space involved in crisis management will benefit from advanced cyber security training that can train multiple different situations, configurations, and attack scenarios without having days of reconfiguration between training days. To provide such training in a condensed time frame, fast automated recovery is crucial. Therefore, we consider our approach for automated recovery as an important step for offering professional training to master real-life cyber attacks.

REFERENCES

1. Candell, R., Zimmerman, T., Stouffer, K.: An industrial control system cybersecurity performance testbed. National Institute of Standards and Technology. NISTIR **8089** (2015)
2. Cybersecurity & Infrastructure Security Agency : Control System Defense: Know the Opponent. Technical report, CISA (2022)
3. Gibadullin, R.F., Nikonorov, V.V.: Development of the system for automated incident management based on open-source software. In: 2021 International Russian Automation Conference (RusAutoCon). pp. 521–525 (2021)
4. Holm, H., Karresand, M., Vidström, A., Westring, E.: A survey of industrial control system testbeds. In: Buchegger, S., Dam, M. (eds.) Secure IT Systems. pp. 11–26. Lecture Notes in Computer Science, Springer International Publishing (2015)
5. Kaiser, F.K., Andris, L.J., Tennig, T.F., Iser, J.M., Wiens, M., Schultmann, F.: Cyber threat intelligence enabled automated attack incident response. In: 2022 3rd International Conference on Next Generation Computing Applications (NextComp). pp. 1–6 (2022). https://doi.org/10.1109/NextComp55567.2022.9932254
6. Lennert, J.F., Retzner, W., Rodgers, M.G., Ruel, B.G., Sundararajan, S., Wolfson, P.D.: The automated backup solution — safeguarding the communications network infrastructure. Bell Labs Technical Journal **9**(2), 59–84 (2004). https://doi.org/10.1002/bltj.20026
7. Masek, P., Stusek, M., Krejci, J., Zeman, K., Pokorny, J., Kudlacek, M.: Unleashing full potential of ansible framework: University labs administration. In: 2018 22nd Conference of Open Innovations Association (FRUCT). pp. 144–150 (2018). https://doi.org/10.23919/FRUCT.2018.8468270
8. Oliveira, D.A.S.d., Crandall, J.R., Wassermann, G., Ye, S., Wu, S.F., Su, Z., Chong, F.T.: Bezoar: Automated virtual machine-based full-system recovery from control-flow hijacking attacks. In: NOMS 2008 - 2008 IEEE Network Operations and Management Symposium. pp. 121–128 (2008). https://doi.org/10.1109/NOMS.2008.4575125
9. President's National Infrastructure Advisory Council (NIAC): Surviving a Catastrophic Power Outage: How to Strengthen the Capabilities of the Nation. Technical report, CISA (2018)
10. Preston, W.C.: Backup and Recovery: Inexpensive Backup Solutions for Open Systems. " O'Reilly Media, Inc." (2007)
11. Swiss Federal Office for Civil Protection: The National Risk Analysis of Disasters and Emergencies in Switzerland. Risk Analysis and Research Coordination (2020)
12. Torkura, K.A., Sukmana, M.I., Cheng, F., Meinel, C.: Slingshot - automated threat detection and incident response in multi cloud storage systems. In: 2019 IEEE 18th International Symposium on Network Computing and Applications (NCA). pp. 1–5 (2019). https://doi.org/10.1109/NCA.2019.8935040
13. Ulrich, J.J., Vaagensmith, B.C., Rieger, C.G., Welch, J.J.: Software defined cyber-physical testbed for analysis of automated cyber responses for power system security. In: 2019 Resilience Week (RWS). vol. 1, pp. 47–54 (2019). https://doi.org/10.1109/RWS47064.2019.8971815
14. Wei, F., Wan, Z., He, H.: Cyber-attack recovery strategy for smart grid based on deep reinforcement learning.

IEEE Transactions on Smart Grid **11**(3), 2476–2486 (2020). https://doi.org/10.1109/TSG.2019.2956161

15. Yun, M., Lan, Y., Han, T.: Automate incident management by decision-making model. In: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). pp. 217–222 (2017). https://doi.org/10.1109/ICBDA.2017.8078811