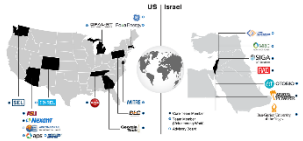
Task 3

Data Collection and Aggregation

OTORIO & Resource Innovations

1/24/2022

The Task Overview

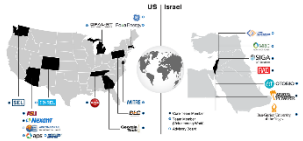


- Task Leader - ASU
- Participants - ASU, BGU, OTORIO, DK Innovation, DLC, Nexant, Delek, Arava
- Task Goal - Lab environments operations, Advisory emulation + Datasets generation

In other words - This is an “Infrastructure project” that its products will be used across the consortium

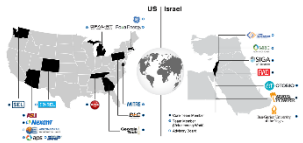
- Task steps -
  - Gather requirements from the different stakeholders
  - Decide on the testing labs - execution plan for each environment
  - At each environment, R&D connectors (plugins)
    - Comprehensive OTORIO - Nexant integration plan in Grid use case
  - Collect, Process and package/expose required data in various forms -
    - Online RAM² interface
    - Offline Datasets

Technical Content



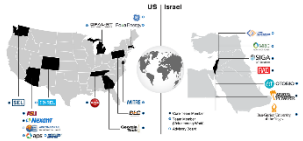
- Testing (lab) environments will be mapped and monitored here
- With each Data provider a systematic process will be taken -
  - Operational Process Use-cases
  - Attack scenarios decision
  - Lab environment model + Data sources enumeration
  - Lab setup, Connectors R&D, Attack scenario R&D
  - Attack execution, Dataset generation & processing

Scenario example



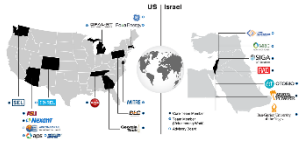
Use Case	1
Process	Refinery
Vendors	Yokogawa, Siemens, Belden Hirschmann
High level architecture	Typical Yokogawa XYZ DCS + DMZ with XY servers + 5 segments… + graph attached
Attack scenario 1	Ransomware infection of OT-DMZ + Control level, infection vector from malicious USB
Attack scenario 2	SIS sabotage - Application configuration change
Attack scenario 3	…
Testing environment	For passive tests - innovation lab in XY with remote access For attack emulation - Export of data + anonymization and testing in Israel's lab
Data sources	PCAPs, Windows event logs, SCADA applications logs + configuration files…

The Team Members and How They Interact



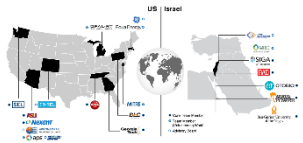
- ASU - **TBD by ASU**
- BGU - State of the art review. Data preparation for benchmarking.
- OTORIO - Lab scenario building, Connectors R&D, Attack scenario R&D, RAM² deployment, Data processing
- Nexant - Grid SME - Lab, attack scenarios, Grid360°-RAM² synergy
- Data providers & Design partners - DK Innovation, DLC, Delek, Arava

Commercialization Plan



- **Datasets commercialization**
  - Multiple IT/OT sources
  - Both raw data and processed
  - Multiple attack scenarios (tagged)
  - Multiple Verticals, processes

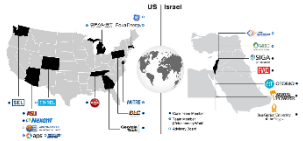
Next Steps



- **Already made tasks**
  - Kickoff meetings with BGU, Delek, Arava
  - Lab Use-cases and attack scenario format, sheet attached
  - Initial review of open ICS security datasets

What	Who	When
Delek - Use cases session	OTORIO, Delek, BGU	TBD
Arava - Use cases session	OTORIO, Arava, BGU	TBD
DLC - Kickoff meeting	OTORIO, DLC, BGU	TBD
Grid Lab	ASU, RI, BGU	TBD

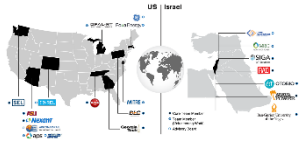
Review of existing ICS datasets



- **Large number of ICS traffic datasets** *Full Packet Capture (FPC) files*
  - 4SICS ICS Lab PCAP files - 360 MB of PCAP files from the ICS village at 4SICS
  - DigitalBond S4x15 ICS Village CTF PCAPs
  - Compilation of ICS PCAP files indexed by protocol (by Jason Smith)
  - PCAP files with OT and IT protocols used in Industrial Control Systems (by ICS Defense / ICS Savunma).
  - DEF CON 23 ICS Village packet captures
  - TRITON execition of the TriStation protocol by Nozomi Networks
  - TriStation traffic
  - Chinese ICS CTF with Modbus/TCP and Siemens S7comm traffic (CTF WP – 工控业务流量分析)
  - ICS Cybersecurity PCAP repository by Univ. of Coimbra CyberSec team

Applicability of the datasets will be evaluated after we gather all internal requirements.

Review of existing ICS datasets



- **OT dataset containing sensor measurements during attack**
  - HIL-based Augmented ICS (HAI) Security Dataset
    - boiler, turbine, water-treatment and HIL simulation
    - 7 datasets 60-229 hours each
    - 88 simulated attack events

time	P1_B2004	P2_B2016	...	P4_HT_LD	attack	attack_P1	...	attack_P3
20190926 13:00:00	0.09830	1.07370	...	0	0	0	...	0

- **dataset containing sensor measurements and command data during attack**
  - gas pipeline dataset and a new version of the same dataset (as AARF) with description

- **dataset with “natural” fault and attack sensor data**
  - power system dataset (description, binary classification, trinary, multiclass)

Datasets with both packet captures and sensor data are missing.