Creation of the UOS\_IOTSH\_2024 Dataset

Our survey of the existing literature at the time of producing this dataset (~March 2024) determined that there is no readily available comprehensive dataset targeting the detection of a sinkhole attack on an RPL-based IoT network covering different network sizes and attack scenarios.

Existing datasets either did not address the specific characteristics of IoT, did not consider the sinkhole attack specifically, or did not consider a broad enough spectrum of network sizes, network topographies, and attack scenarios. Therefore, we embarked on the creation the UOS-IOTSH\_2024 Dataset that should be comprehensive enough to support future research work on detection of sinkhole attacks. The following subsections will describe the development of the UOS\_IOTSH\_2024 dataset.

# Attack Scenario

The UOS\_IOTSH\_2024 Dataset simulates an internal attack scenario, where an intruder infiltrates a node that is already integrated into a functioning RPL-based IoT network to execute a sinkhole attack. This scenario mimics a particularly dangerous form of attack, where an adversary leverages a compromised node's established network privileges to orchestrate malicious activities. By compromising an existing node within the network, the attacker gains the ability to covertly manipulate the network's communication dynamics, directing traffic flow in a manner that disrupts normal operations and compromises network integrity.

Once the intruder has gained access to the victim node, it will modify the rank information advertised by the compromised node. By falsely advertising an optimal path to the root node, the attacker can attract traffic towards the compromised node, effectively hijacking the network's communication pathways. Because the attack originates from within the network, it can be very challenging to detect using traditional security mechanisms.

# Requirements for the UOS\_IOTSH\_2024 Dataset

The UOS\_IOTSH\_2024 Dataset has been developed to meet the following requirements:

1. **Network Topologies.** The dataset should encompass a variety of network topologies, including small-scale and medium-scale networks, to represent different deployment scenarios commonly encountered in RPL-based IoT networks.
2. **DODAGs.** The dataset should include networks with varying numbers of DODAGS to investigate the impact of multiple DODAGs on detection accuracy and latency.
3. **Sinkhole Attack Scenarios.** The dataset should represent different sinkhole attack scenarios, including single sinkhole attacks and double sinkhole attacks, with varying degrees of severity and impact on network behavior.
4. **Attacker Locations.** The dataset should include attackers in different locations within the hierarchy to enable assessing the effect of the attacker's location on detection accuracy.
5. **Anomalous Patterns.** The dataset should include instances of normal network operation as well as network operation under sinkhole attack to facilitate the development and evaluation of ML-based detection methods.
6. **Realistic Network Traffic.** The dataset should incorporate realistic network traffic patterns, including both control and data packets, to accurately simulate the behavior of RPL-based IoT networks.
7. **Non-Invasive Detection.** The dataset should be based on the standard control packets generated by the RPL protocol without any modifications to the protocol.
8. **Ground Truth Labels.** The dataset should be annotated with ground truth labels indicating the presence or absence of sinkhole attacks in each scenario, allowing for supervised learning approaches in ML/DL-based detection methods.

These requirements have been used to guide the development of the dataset as will be shown in the coming subsections.

# Composition of the UOS\_IOTSH\_2024 Dataset

Based on the above requirements, the dataset was created with the following main categories:

1. Traffic from RPL-based IoT networks operating normally under multiple scenarios:
   1. A small network with a single DODAG.
   2. A medium-sized network with a single DODAG.
   3. A medium-sized network with dual DODAGs.
2. Traffic from RPL-based IoT networks under a single sinkhole attack:
   1. A small network with a single DODAG
      * multiple instances for different attacker locations
   2. A medium-sized network with a single DODAG
      * multiple instances for different attacker locations
   3. A medium-sized network with dual DODAGs
      * multiple instances for different attacker locations
3. Traffic from RPL-based IoT networks under dual sinkhole attacks:
   1. A small network with a single DODAG
      * multiple instances for different attacker locations
   2. A medium-sized network with a single DODAG
      * multiple instances for different attacker locations
   3. A medium-sized network with dual DODAGs
      * multiple instances for different attacker locations

# The Attack Effect Numerical Measure

Under certain circumstances it becomes very difficult and time consuming to exhaustively investigate all possible attack scenarios on a particular setup. This is especially true when multiple attackers may be present in the network and the possible combinations of scenarios becomes prohibitively large. Under these circumstances it pays to concentrate the effort on the worst-case scenarios which would cause the worst damage to the network traffic.

We created a numerical measure to assess the effect of any particular attack on the exchanged traffic in the network. The *Sinkhole* *Attack Effect* measure reflects the attack's success in redirecting network traffic toward the compromised node(s). Essentially, it estimates the percentage of total network traffic routed through the compromised node(s) following the attack. A higher percentage indicates a more severe impact of the attack on the network's normal operation. The *Sinkhole* *Attack Effect* measure is calculated according to the expression in equation (1):

|  |  |
| --- | --- |
|  | (1) |

where “Affected Nodes” refers to nodes whose traffic passes directly or indirectly through the compromised node(s) after the attack.

# Determination of Appropriate Network Sizes to Simulate

Analysis of the literature available at the time of creating this dataset (~March 2024) has shown that previous researchers have used three ranges of network sizes as shown in Table 1.

Table 1 – Network Size Ranges Employed in the Literature

|  |  |  |
| --- | --- | --- |
| Network Size | Number of nodes | References |
| Small | Less than 20 | [1], [2], [3], [4], [5], [6], [7], [8], and [9] |
| Medium | Less than 50 | [1], [5], [10], [11], and [12] |
| Large | More than 100 | [13], [11], [14], [15], [16], [10], [17], [18], and [19] |

Accordingly, it was decided to adopt a network size of 12 sensor nodes (in addition to the root node) to represent a small network and a network size of 24 sensor nodes (in addition to one or two root nodes) to represent the medium-sized network.

# Features Included in the UOS\_IOTSH\_2024 Dataset

The COOJA simulator allows access to a large number of features and parameters from the RPL packets. However, not all of these features are used during the crucial period of constructing the DODAG. Additionally, most of these features are not affected by the nature of the sinkhole attack. Given that the UOS\_IOTSH\_2024 dataset is targeted specifically at sinkhole attack scenarios, we chose to include only a small collection of fields in the dataset to keep the number of features manageable and concentrate on the fields that may be affected by the sinkhole attack. We considered the following fields for inclusion:

* All the fields in the RPL control messages that are used for constructing and maintaining the DODAG and controlling the operation of the network.
* The DODAG configuration options which are configured in the root node and distributed to the rest of the DODAG nodes.
* Certain details obtained from the IPv6 packet header, such as the source and destination IP addresses.

Then, these parameters were filtered by assessing the impact of the attack on each specific feature. Accordingly, the dataset was first created with the 6 features shown in Table 2 extracted directly from the raw network traffic collected from the COOJA simulations.

Table 2 – RPL Fields Included in the UOS\_IOTSH\_2024 Dataset

| Field | Feature Description | RPL Control Message | Reason of inclusion or omission |
| --- | --- | --- | --- |
| Source ID | IPv6 address of the source node | IPv6 header | Used to identify child and parent nodes, which are affected by the sinkhole attack |
| Destination ID | IPv6 address of the destination node | IPv6 header |
| Rank | Relative position of node relative to other nodes with respect to a DODAG root. | DIO | This is the exact attack target of sinkhole attack. |
| Length | The length of the RPL control message (64: DIS, 97 or 102: DIO, 76: DAO or 5: DAO ACK). | DODAG Configuration Option | The sequence of the control messages between the nodes is affected by the sinkhole attack. |
| Info | Indicate the type of the RPL control message | DODAG Configuration Option | To indicate the type of the control message. The attack utilized DIO message to announce the false rank, and the DAO message is affected by the attack existence due to the re-election of the parent. |
| Protocol | The communication protocol (ICMPV6, IEEE802.1 or UDP). | DODAG Configuration Option | To simulate the real network traffic |

# Procedure for Simulating the Attack using the COOJA Simulator

The procedure for simulating the internal sinkhole attack described in the attack scenario of section ‎1 proceeded according to the following steps:

1. The chosen attack scenario assumes an internal attack leveraging a node that is already integrated into an existing DODAG. Therefore, after the network is constructed in COOJA, it will be allowed to run normally until the DODAG is constructed successfully and the network starts to transmit UDP data packets. Through preliminary experimentation, we determined that this process typically requires approximately 100 seconds of simulation time. Accordingly, the network will be allowed to run normally for 100 seconds to achieve convergence and stability, then the attack will be launched.
2. After the conclusion of the 100 seconds, a designated node will alter its behavior and launch the sinkhole attack by broadcasting DIO messages that falsely advertise a rank very close to that of the root node. The location of this compromised node will be varied to investigate the impact of the attacker location on the detection accuracy.
3. The simulator will proceed to implement normal RPL behavior, where nodes continuously update their adopted parents by choosing nodes within their communication range with the lowest advertised rank, assuming they are closer to the root. Since the compromised node advertises a rank very close to that of the root node, neighboring nodes will perceive it as offering a shorter path to the root, gradually adopt it as their preferred parent and reroute their traffic accordingly. This marks a successful launch of the sinkhole attack.
4. Once the network reconstruction process is finished, the network transitions to the data packet transmission phase. Extensive trial and error experiments have demonstrated that approximately 4 minutes of simulation time (counted from the beginning of step 1) are sufficient to reach this stage. Given that the focus is the detection of the attack using exchanged control packets, further simulation beyond the initiation of data packet transmission is deemed unnecessary. Therefore, it was decided that the simulations would be allowed to run for only 4 minutes.
5. After the conclusion of the 4 minutes, Wireshark will be used to collect all of the exchanged packets throughout the simulation time. These packets will be recorded in the Dataset, with their exact sequencing preserved.

# Collection of Normal Network Traffic for the UOS\_IOTSH\_2024 Dataset

A number of network configurations were built in COOJA and allowed to operate normally for 4 minutes (as established in section ‎7) to allow the DODAG construction process to complete and converge. The network traffic under these circumstances forms the baseline *normal behavior* for the simulated networks. The following subsections will detail the process of collecting the data for each scenario.

## Small Single-DODAG Network

A network comprised of a single router (representing the root node) and 12 Tmote Sky Mote nodes (representing the sensor nodes) was constructed in the COOJA simulator. The sensor nodes were arranged in rows with varying distances from the root (as illustrated in Figure 1) to encourage a hierarchical DODAG. To prevent potential data bias, the placement of nodes with respect to their node IDs was deliberately randomized.

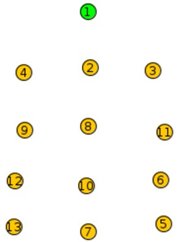


Figure 1 – Arrangement of Nodes to Create the Small Single-DODAG Network in COOJA.

The simulator was then operated continuously for a duration of 4 minutes of simulation time. After that, Wireshark was used to capture all ICMP4 and UDP packets exchanged during the simulation period. Careful examination of the information in the collected packets showed that the network assembled appropriately into the DODAG shown in Figure 2. The distribution of the collected packets during the 4-minute simulation time is presented in Table 3.

A diagram of a network

Description automatically generated

Figure 2 – DODAG of the Small Network

Table 3 –Packet Distribution in the Small Single-DODAG Network under Normal Behavior.

|  |  |  |
| --- | --- | --- |
| Packet Type | Count | Percentage |
| ICMPv6 – DIS | 228 | 4% |
| ICMPv6 – DIO | 2,662 | 45% |
| ICMPv6 – DAO | 2,089 | 35% |
| ICMPv6 – DAO ACK | 239 | 4% |
| UDP | 705 | 12% |
| **Totals** | **5,923** | **100.00%** |

## Medium-sized Single-DODAG Network

A medium-sized network comprised of a single root node and 24 Tmote Sky Mote nodes was constructed in COOJA. The sensor nodes were distributed in a manner similar to the previous network to encourage a hierarchical construction of the DODAG (as demonstrated in the diagram of Figure 3). Then, the same steps were followed for collecting the network traffic from the COOJA simulator. The statistics about the collected data is summarized in Table 4.

A group of yellow circles with green and black numbers

Description automatically generated

Figure 3 – Arrangement of Nodes to Create the Medium-sized Single-DODAG Network.

Table 4 –Packet Distribution in the Medium-sized Single-DODAG Network under Normal Behavior.

|  |  |  |
| --- | --- | --- |
| Packet Type | Count | Percentage |
| ICMPv6 – DIS | 456 | 1% |
| ICMPv6 – DIO | 6,998 | 19% |
| ICMPv6 – DAO | 24,571 | 67% |
| ICMPv6 – DAO ACK | 819 | 2% |
| UDP | 3,915 | 11% |
| **Totals** | **36,759** | **100.00%** |

## Medium-sized Dual-DODAG Network under Normal Operation

In order to facilitate investigating the behavior of a sinkhole attack on a network with multiple DODAGs, we chose to incorporate samples from a dual-DODAG network into the dataset. Accordingly, a medium-sized network comprised of two routers (representing two root nodes) and 24 Tmote Sky Mote nodes was built in COOJA. The nodes were distributed in a manner to encourage the establishment of two DODAGs.

The simulator was allowed to run normally for the 4 minutes, resulting in the nodes automatically organizing into the two DODAGs illustrated in Figure 4. Each DODAG contained a single root node (node 1 and 14 respectively) and 12 sensor nodes. Statistics about the collected traffic is presented in Table 5.

A diagram of a network

Description automatically generated with medium confidence

Figure 4 – Topology of a Medium-sized Dual-DODAG Network

Table 5 –Packet Distribution in the Medium-sized Dual-DODAG Network under Normal Behavior.

|  |  |  |
| --- | --- | --- |
| Packet Type | Count | Percentage |
| ICMPv6 – DIS | 450 | 3% |
| ICMPv6 – DIO | 6,076 | 36% |
| ICMPv6 – DAO | 7,588 | 45% |
| ICMPv6 – DAO ACK | 544 | 3% |
| UDP | 2,041 | 12% |
| **Totals** | **16,699** | **100%** |

# Collection of Traffic from Networks Subjected to a Single Attacker

The same three combinations of network sizes and structures from the previous section were subjected to a single sinkhole attack to enable the extraction of representative traffic from networks under attack.

## Small Single-DODAG Network under a Single Sinkhole Attack

The network of Figure 1 was recreated exactly. Then, the procedure described in section ‎7 was followed to simulate compromising node number 2 to launch an internal sinkhole attack on the network at the 100 second mark. Examination of the collected messages reveals that the nodes initially established the same DODAG as in section ‎8.1 (depicted in the “before” diagram of Figure 5). However, following the launch of the attack, the DODAG structure underwent significant modification, as depicted in the “after” diagram in the figure.

|  |  |
| --- | --- |
|  |  |
| Before | after |

Figure 5 – A Small Single-DODAG Network before and after a Sinkhole Attack at Node 2

To ensure the dataset encompasses a broad range of scenarios with attackers originating from various points within the network, we conducted an exhaustive simulation where a different node was selected as the compromised node for each iteration. As a result, 12 different instances of the *Small Single-DODAG Network* segment were added under the *Single Attacker* category of the dataset. The statistics for the network traffic collected during the 4-minute simulation time are presented in Table 6. The values in the table show the counts for one representative instance out of the 12.

Table 6 – Packet Distribution in the Small Single-DODAG Network under a Single Sinkhole Attack.

|  |  |  |
| --- | --- | --- |
| Packet Type | Count | Percentage |
| ICMPv6 – DIS | 228 | 4% |
| ICMPv6 – DIO | 2,583 | 46% |
| ICMPv6 – DAO | 1,884 | 34% |
| ICMPv6 – DAO ACK | 235 | 4% |
| UDP | 652 | 12% |
| **Totals** | **5,582** | **100%** |

## Medium-sized Single-DODAG Network under a Single Sinkhole Attack

The medium-sized single-DODAG network of Figure 3 was recreated and simulated with one of the nodes launching a sinkhole attack after 100 seconds of simulation time. Similarly, the simulation was run exhaustively with a different node being selected as the attacking node for each iteration. Twenty-four different instances representing the attack being launched from each of the 24 nodes were added to the dataset. The distribution of packets collected during the 4-minute simulation time is presented in Table 7. The values in the table show the counts for one representative instance out of the 24.

Table 7 – Packet Distribution in the Medium-sized Single-DODAG Network under a Single Sinkhole Attack.

|  |  |  |
| --- | --- | --- |
| Packet Type | Count | Percentage |
| ICMPv6 – DIS | 532 | 1% |
| ICMPv6 – DIO | 8566 | 23% |
| ICMPv6 – DAO | 23005 | 62% |
| ICMPv6 – DAO ACK | 1002 | 3% |
| UDP | 4237 | 11% |
| **Totals** | **37,342** | **100%** |

## Medium-sized Dual-DODAG Network under a Single Sinkhole Attack

A dual-DODAG network essentially consists of two independent small single-DODAG sub-networks. When a single attack is launched on such a network, one DODAG will continue to operate normally, similar to the scenario discussed in section ‎8.1, while the other DODAG behaves as described in section ‎9.1. In such cases, there is nothing particularly noteworthy about the samples extracted under these conditions.

On the other hand, if the attack is initiated from a node on the border between the two DODAGs, significant variation may occur in the traffic patterns. Here the attack may alter the distribution of the nodes between the two DODAGs and attract traffic from nodes that were previously in the other DODAG, as illustrated in Figure 6.

|  |  |
| --- | --- |
|  |  |
| DODAG Structures under Normal Behavior | Effect of Sinkhole Attack on Node 6 |

Figure 6 – A Medium-sized Dual-DODAG Network with a Single Attack Targeting a Node on the Border between the DODAGs

Accordingly, the network of Figure 4 was recreated and simulated with an attack being launched using one of the border nodes after 100 seconds of simulation time. The simulation was repeated exhaustively for each of the border nodes. This resulted in eight different instances being added to the dataset representing the traffic streams collected from each border node in the network. The distribution of the packets collected during the 4-minute simulation time is shown in Table 8. The values in the table show the counts for one representative instance out of the 8 instances.

Table 8 – Packet Distribution in the Medium-sized Dual-DODAG Network under a Single Sinkhole Attack.

|  |  |  |
| --- | --- | --- |
| Packet Type | Count | Percentage |
| ICMPv6 – DIS | 456 | 3% |
| ICMPv6 – DIO | 5,731 | 34% |
| ICMPv6 – DAO | 8,291 | 50% |
| ICMPv6 – DAO ACK | 566 | 3% |
| UDP | 1,699 | 10% |
| **Totals** | **16,743** | **100.00%** |

# Collection of Traffic from Networks Subjected to Dual Attackers

A network may be infected by multiple attackers at the same time. Therefore, we created a category in the dataset to represent scenarios where the network is attacked by two attackers. It is highly unlikely that the two attacks will be launched simultaneously. Therefore, our simulations consider that one of the nodes launches an attack at the 100 second mark, while the second node launches the attack 50 seconds later.

The same network structures described in the previous two sections were subjected to two attackers to enable extracting the representative samples for the dataset.

## Small Single-DODAG Network under a Dual Sinkhole Attack

The network of Figure 2 was recreated and simulated with two nodes launching sinkhole attacks. Rather than attempting an exhaustive simulation of all the possible combinations of two attacking nodes in a network of 12 nodes (132 combinations), we used the *Sinkhole* *Attack Effect* measure to evaluate different scenarios involving two attackers at different location in the network to determine the combinations producing the worst-case effect.

In the first step, an exhaustive analysis was conducted for the sinkhole attack effect of a single attack at each of the nodes in the network (as documented in Table 9). These scores were used to gauge the correlation between the topological level of the attacker and the effect of an attack from that level. The results in the table show that an attack from level 1 would produce the worst effect due to the proximity to the root nodes.

Table 9 – Demonstration of Attack Effect on Network Traffic.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Topological Level** | **Malicious Node ID** | **Nodes** | | **# of affected nodes** | **Sinkhole Attack Effect** | **Level effect “Trend”** |
| **Affected** | **Not affected** |
| L1 | 2 | 6,8,10,5,8 ,11,13 | 4,12,3 | 7 | 64% | 61% |
| 3 | 6,8,5,7,11,13 | 2,4,10,12,9 | 6 | 55% |
| 4 | 8,10,12,5,7,9,13 | 2,3,6,11 | 7 | 64% |
| L2 | 6 | 8,10,5,13 | 2,3,4,12,7,9,11 | 4 | 36% | 38% |
| 8 | 5,7,9,11,13 | 2,3,4,6,10,12 | 5 | 45% |
| 9 | 12,7,13 | 2,3,4,5,6,8,10,11 | 3 | 27% |
| 10 | 6,8 ,5,11 | 2,3,4,7,9,12,13 | 4 | 36% |
| 11 | 6,5,7 | 2,3,4,8,10,12,9 ,13 | 3 | 27% |
| 12 | 8,10,5,7,9,13 | 2,3,4,6,11 | 6 | 55% |
| L3 | 5 | 7,12 | 2,3,4,6,8,10,9,11,13 | 2 | 18% | 45% |
| 7 | 8,10,12,5,13 | 2,3,4,6,9,11 | 5 | 45% |
| 13 | 8,12,5,7,9 | 2,3,4,6,10,11 | 5 | 45% |

It was noticed that the attack effect dropped significantly when the attacking node was in level 2 compared to when it was in level 1. On the other hand, the attack effect increased with the increase in the topological level of the attacker. This can be attributed to the fact that nodes in level 1 will be little affected by an attacker at a more distant level, and will maintain their parent selection integrity. Consequently, nodes at level 1 remain unharmed if the attacker is positioned at, for instance, level 2. However, as we move to levels further from the root, nodes positioned beyond the transmission range of the root become more susceptible to the influence of the attacker.

Based on these evaluations, it is determined that having two attackers in level 1 of the DODAG would be totally catastrophic to the network and would not offer any insight on the behavior of the IDS under the attack scenario. Therefore, this combination was not included in favor of more interesting and challenging combinations.

On the other hand, Table 10 demonstrates that certain combinations of attack locations produced significant combined sinkhole attack effect scenarios. Therefore, based on these considerations, we decided that we would include only the following representative combinations in the UOS\_IOTSH\_2024 Dataset:

* One attacker at L2 and another attacker at L3: Node 8 and Node 2
* One attacker at L2 and another attacker at L3: Node 11 and Node 12
* One attacker at L2 and another attacker at L4: Node 5 and Node 9
* One attacker at L3 and another attacker at L4: Node 5 and Node 12

Simulations were carried out for these scenarios and the network traffic was collected in a manner similar to what was discussed in earlier steps. Four instances were added to the dataset to represent these four scenarios. The distribution of the collected samples for these scenarios is shown in Table 11. The values in the table show the counts for one representative instance out of the 4 instances.

Table 10 – Samples of Sinkhole Attack Effect in Dual-Attacker Scenario

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Level** | **Compromised Node ID** | **Nodes** | | **# of affected nodes** | **Sinkhole Attack Effect** |
| **Affected** | **Not affected** |
| 1st attack at level 2 | 8 (after 100 sec) | 6, 10, 7, 9, 11, 13 | 4, 12, 3, 5 | 6 | 60% |
| 2nd attack at level 1 | 2 (after 150 sec) |
| 1st attack at level 3 | 5 (after 100 sec) | 7, 12, 6 | 10, 3,2,4,8,11, 13 | 3 | 30% |
| 2nd attack at level 2 | 9 (after 150 sec) |
| 1st attack at level 3 | 5 (after 100 sec) | 7, 13, 6 | 2,3,4,10,11,8,9 | 3 | 30% |
| 2nd attack at level 2 | 12 (after 150 sec) |
| 1st attack at level 2 | 11 (after 100 sec) | 13 | 2,3,4,8,10,9,7,6,5 | 1 | 10% |
| 2nd attack at level 2 | 12 (after 150 sec) |

Table 11 – Packet Distribution in the Small Single-DODAG Network under Dual Sinkhole Attacks.

|  |  |  |
| --- | --- | --- |
| Packet Type | Count | Percentage |
| ICMPv6 – DIS | 228 | 4% |
| ICMPv6 – DIO | 2,593 | 45% |
| ICMPv6 – DAO | 2,025 | 35% |
| ICMPv6 – DAO ACK | 226 | 4% |
| UDP | 695 | 12% |
| **Totals** | **5,767** | **100.00%** |

## Medium-sized Single-DODAG Network under a Dual Sinkhole Attack

The network of Figure 3 was recreated and simulated with two nodes launching attacks with a 50 second interval. The attack effect score was employed to evaluate different scenarios involving two attackers in diverse network positions to determine the combinations producing the most severe attack effects. A sample of these results is shown in Table 12.

Simulations were carried out for the scenarios in the table and the network traffic was collected in a manner similar to what was discussed in earlier steps. Five components were added to the dataset to represent these five scenarios. The distribution of the collected samples for a representative sample of these scenarios is shown in Table 13.

Table 12 – Attack Effect of Different Scenarios Involving Two Attackers in a Medium-sized Single-DODAG network.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Compromised Nodes ID | Nodes | | # of affected nodes | % of compromised network traffic |
| Affected | Not affected |
| 15 (after 100 sec) | 3,4,5,7,8,9,11,12,13,16, 18,19,20,21,22,23,24,26 | 6,10,17,25 | 18 | 82% |
| 2 (after 150 sec) |
| 3 (after 100 sec) | 2,4,5,7,8,9,11,12,13,16,18,19,20,21,22,26 | 6,10,15,17,23,24 | 16 | 73% |
| 25 (after 150 sec) |
| 18 (after 100 sec) | 3,4,5,7,8,9,10,11,12,16,19,20,21,22,23 | 2,6,15,17,24,25,26 | 15 | 68% |
| 13 (after 150 sec) |
| 16 (after 100 sec) | 2,4,5,7,8,11,12,13,18,19, 21,26 | 3,6,10,15,17,20,22,23,24, 25 | 13 | 59% |
| 9 (after 150 sec) |
| 12 (after 100 sec) | 4,7,8,11,13,18,19,21,22 | 2,3,6,9,10,15,16,17,20,23,24,25,26 | 9 | 41% |
| 5 (after 150 sec) |

Table 13 – Packet Distribution in the Medium Sized Single-DODAG Network under Dual Sinkhole Attacks.

|  |  |  |
| --- | --- | --- |
| Packet Type | Count | Percentage |
| ICMPv6 – DIS | 646 | 2% |
| ICMPv6 – DIO | 8,196 | 22.5% |
| ICMPv6 – DAO | 21,382 | 59% |
| ICMPv6 – DAO ACK | 926 | 2.5% |
| UDP | 5,223 | 14% |
| **Totals** | **36,373** | **100.00%** |

## Medium-sized Dual-DODAG Network under a Dual Sinkhole Attack

The network of Figure 4 was recreated and simulated with two nodes launching a sinkhole attack. Given that this network topology is a combination of two small networks, therefore, investigating the situation of both attacks being launched in the same DODAG will devolve into the same behavior investigated in section ‎10.1. Therefore, the dual attack simulations for this setup were implemented by initiating each attack in a separate DODAG.

The following four representative combinations were selected through informed guessing:

* Node 7 (DODAG 1) and Node 26 (DODAG 2)
* Node 17 (DODAG 2) and Node 12 (DODAG 1)
* Node 8 (DODAG 1) and Node 19 (DODAG 2)
* Node 10 (DODAG 1) and Node 20 (DODAG 2)

Then the network traffic was collected in a manner similar to what was discussed in earlier steps. Four components were added to the dataset to represent these four scenarios. The distribution of the collected samples from a representative instance is shown in Table 14.

Table 14 – Packet Distribution in the Medium-Sized Dual-DODAG Network under Dual Sinkhole Attacks.

|  |  |  |
| --- | --- | --- |
| Packet Type | Count | Percentage |
| ICMPv6 – DIS | 456 | 3% |
| ICMPv6 – DIO | 5,639 | 32% |
| ICMPv6 – DAO | 8,884 | 51% |
| ICMPv6 – DAO ACK | 595 | 3% |
| UDP | 1,874 | 11% |
| **Totals** | **17,448** | **100.00%** |

# Complete composition of the UOS\_IOTSH\_2024 Dataset

Once the samples for all of the different scenarios were collected, we ended up with a dataset comprised of 9 different categories covering different network sizes, different network topologies, and different attacker counts. In total, the dataset covers 57 different attack scenarios, and includes more than one million samples of network traffic as shown in Table 15.

Table 15 – Composition of the UOS\_IOTSH\_2024 Dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Component | Number of Scenarios per Component | Number of Samples per Scenario | Number of Samples per Component |
| Small Single-DODAG Network under Normal Behavior | 1 | 5,923 | 5,923 |
| Medium-sized Single-DODAG Network under Normal Behavior | 1 | 36,759 | 36,759 |
| Medium-sized Dual-DODAG Network under Normal Behavior | 1 | 16,699 | 16,699 |
| Small Single-DODAG Network under a Single Sinkhole Attack | 12 | 5,582 | 66,984 |
| Medium-sized Single-DODAG Network under a Single Sinkhole Attack | 24 | 37,342 | 896,208 |
| Medium-sized Dual-DODAG Network under a Single Sinkhole Attack | 8 | 16,743 | 133,944 |
| Small Single-DODAG Network under Dual Sinkhole Attacks | 4 | 5,767 | 23,068 |
| Medium Sized Single-RPL Network under Dual Sinkhole Attacks | 5 | 36,373 | 181,865 |
| Medium Sized Dual-DODAG Network under Dual Sinkhole Attacks | 4 | 17,449 | 69,796 |
| **Total** | **60** |  | **1,431,246** |

# Data Preprocessing for the UOS\_IOTSH\_2024 Dataset

Some of the features mentioned in Table 2 required some preparatory modifications before they can be used in training of ML and DL models. Consequently, the following pre-processing steps were applied to all extracted packets before being included in the dataset:

1. Conversion of the hexadecimal values in the “source IP address” and “destination IP address” columns using one-hot encoding to enable the ML/DL models to learn distinct patterns linked to each IP address.
2. Encoding “Info” column which displays the message type (DIO, DAO, DIS, and ACK) in string format. The same step is also performed to the “Protocol” column that determines the protocol type (ICMPv6, UDP, and IEEE 802.15.4). The different string values were encoded to enable the ML and DL models to understand their values.

These steps did not affect the content of the fields, but rather changed the representation of the data into a format better understandable by the modeling environment.

# Sample Labelling for the UOS\_IOTSH\_2024 Dataset

The sinkhole attack disrupts the network topology causing effects possibly reaching very far across the network. Accordingly, it is possible that the entire network traffic may be affected after the attack. Therefore, one should consider the traffic of the entire network while trying to detect the attack. Based on this understanding, a labelling approach based on the status of the entire network rather than a specific node was chosen. A ‘*network operating normally*’ label was applied to all of the traffic in the dataset ‘*before*’ the attack is launched, and a ‘*network under attack*’ label to all of the traffic ‘*after*’ the attack is launched. This labelling approach will enable the detection of anomalies in the network behavior caused by the attack, but will not enable identifying the malicious node specifically.

The sample distribution illustrated in Figure 7 was achieved. The distribution was slightly unbalanced favoring ‘*normal*’ samples. This can be attributed to the fact that more traffic is exchanged during the early stages of the DODAG construction (the first 100 seconds) compared to the amount exchanged during the DODAG update once the attack is launched (after the first 100 seconds).

Figure 7 – Sample Distribution in the Collected Dataset.

# References

[1] Raza S, Wallgren L and Voigt T 2013 SVELTE: Real-time intrusion detection in the Internet of Things *Ad Hoc Networks* **11** 2661–74

[2] Mirshahjafari S M H and Ghahfarokhi B S 2019 Sinkhole+CloneID: A hybrid attack on RPL performance and detection method *Inf. Secur. J.* **28** 107–19

[3] Bostani H and Sheikhan M 2017 Hybrid of anomaly-based and specification-based IDS for Internet of Things using unsupervised OPF based on MapReduce approach *Comput. Commun.* **98** 52–71

[4] Alzubaidi M, Anbar M, Chong Y W and Al-Sarawi S 2018 Hybrid monitoring technique for detecting abnormal behaviour in rpl-based network *J. Commun.* **13** 198–208

[5] Taghanaki S R, Jamshidi K and Bohlooli A 2019 DEEM: A decentralized and energy efficient method for detecting sinkhole attacks on the internet of things *2019 9th Int. Conf. Comput. Knowl. Eng. ICCKE 2019* **2019** 325–30

[6] Yang W, Wang Y, Lai Z, Wan Y and Cheng Z 2019 Security Vulnerabilities and Countermeasures in the RPL-Based Internet of Things *Proceedings - 2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, CyberC 2018* (Conf. Cyber-Enabled Distrib. Comput. Knowl. Discov. CyberC 2018) pp 49–54

[7] Muzammal S M, Murugesan R K, Jhanjhi N Z, Humayun M, Ibrahim A O and Abdelmaboud A 2022 A Trust-Based Model for Secure Routing against RPL Attacks in Internet of Things *Sensors* **22**

[8] Airehrour D, Gutierrez J A and Ray S K 2019 SecTrust-RPL: A secure trust-aware RPL routing protocol for Internet of Things *Futur. Gener. Comput. Syst.* **93** 860–76

[9] Napiah M N, Bin Idris M Y I, Ramli R and Ahmedy I 2018 Compression Header Analyzer Intrusion Detection System (CHA - IDS) for 6LoWPAN Communication Protocol *IEEE Access* **6** 16623–38

[10] Patel B and Shah P 2021 Direct Neighbour Sink Reputed Trust Based Intrusion Detection System to Mitigate Sinkhole Attack in RPL for IoT Networks *J. Eng. Sci. Technol. Rev.* **14** 38–45

[11] Cervantes C, Poplade D, Nogueira M and Santos A 2015 Detection of sinkhole attacks for supporting secure routing on 6LoWPAN for Internet of Things *Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management, IM 2015* (Symp. Integr. Netw. Manag. IM 2015) pp 606–11

[12] Tabari M Y and Mataji Z 2020 Detecting Sinkhole Attack in RPL-based Internet of Things Routing Protocol *J. AI Data …* **9** 73–85

[13] Surendar M and Umamakeswari A 2016 InDReS: An Intrusion Detection and response system for Internet of Things with 6LoWPAN *Proceedings of the 2016 IEEE International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2016* (WiSPNET 2016: Conf. Wirel. Commun. Signal Process. Networking) pp 1903–8

[14] Zaminkar M, Sarkohaki F and Fotohi R 2021 A method based on encryption and node rating for securing the RPL protocol communications in the IoT ecosystem *Int. J. Commun. Syst.* **34** 1–18

[15] Zaminkar M and Fotohi R 2020 SoS-RPL: Securing Internet of Things Against Sinkhole Attack Using RPL Protocol-Based Node Rating and Ranking Mechanism *Wirel. Pers. Commun.* **114** 1287–312

[16] Choudhary S and Kesswani N 2019 Cluster-based intrusion detection method for internet of things *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA* vol 2019-Novem (vol. 2019-Novem: Conf. Comput. Syst. Appl. AICCSA) pp 1–8

[17] Farooq U, Asim M, Tariq N, Baker T and Awad A I 2022 Multi-Mobile Agent Trust Framework for Mitigating Internal Attacks and Augmenting RPL Security *Sensors* **22**

[18] Prathapchandran K and Janani T 2021 A trust aware security mechanism to detect sinkhole attack in RPL-based IoT environment using random forest – RFTRUST *Comput. Networks* **198**

[19] Kamel S O M and Elhamayed S A 2020 Mitigating the impact of iot routing attacks on power consumption in iot healthcare environment using convolutional neural network *Int. J. Comput. Netw. Inf. Secur.* **12** 11–29