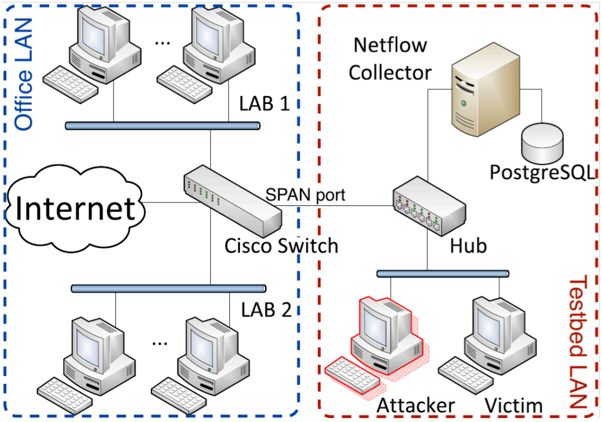
Port Scanning attack dataset

Description of how data were collected

The data traffic has been gathered from a Local Area Network (LAN) in a research office environment, in the Wolfson School at Loughborough University. The PCs in two distinct labs are connected to the same office LAN, and these PCs are used by researchers daily, mostly for Internet access. In addition, two additional PCs have been connected through a hub to a testbed LAN in order to implement the **port scanning attacks**. An attacker runs Linux Ubuntu and launches the attack using the network mapping tool Nmap, and a victim that also runs Linux Ubuntu and is also in charge of gathering all the network traffic used in our experiments. The victim PC was also used to access the Internet during the experiments.

Network traffic from all the PCs in the office LAN is mirrored and routed in the SPAN port of a switch. The switch aggregates the traffic from all the PCs in the office LAN using the SPAN port. The aggregated traffic is carried over to the testbed LAN through the hub connected to the switch. The network data traffic generated by the PCs in the office LAN is used in our experiments as realistic background network traffic. The figure below shows the logical topology of the testbed LAN. The left part of the figure represents the PCs that have generated the background network traffic, whereas the right part represents the Netflow collector and the PCs (attacker and victim) involved in the port scanning attacks.

In this experiment, we have focused on vertical port scan only. In the vertical scan, either all the ports or a range of ports are scanned in one targeted host.



Logical topology of the testbed LAN; PCs on the left generate the background traffic, while those on the right are involved in the port scanning attacks implementation, and detection process.

Description of raw data

All the network traffic from the testbed and office LANs has been collected and gathered by the victim using the network analyser Tcpdump in pcap format. In total, 160 GBytes of network traffic have been gathered during the 9 days that the experiment lasted (from 21st July to 29th July, 2016). The pcap traffic dataset comprises 99.40% of non- malicious traffic (i.e. 696638 data instances) and 0.60% of malicious traffic (i.e. 4220 data instances).

Description of provided data

The provided file includes the following columns:

[1] Date dd/MM/YY
[2] Time hh:mm:ss
[3] Number of frames transmitted per second
[4] Number of distinct Source MAC addresses per second
[5] Number of distinct Destination MAC addresses per second
[6] Throughput (bytes per second)
[7] Number of distinct Source Ports per second
[8] Number of distinct Destination Ports per second
[9] Flag (i.e., 1 if entry includes port scanning activity or 2, if it is just normal traffic)

Each row is a per second instance of aggregated metric measurements as seen above.