A Flow-based Entropy Characterization of a NATed Network and its Application on Intrusion Detection

Elie Kfoury, Jose Gomez

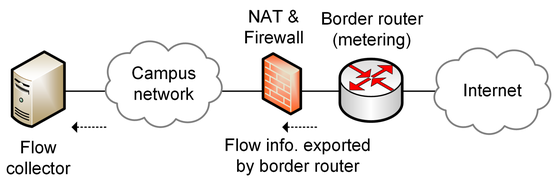
Integrated Information Technology Department, University of South Carolina, Columbia, South Carolina

Abstract

- This project presents a flow-based entropy characterization of a small/medium-sized campus network that uses network address translation (NAT).
- Although most networks follow this configuration, their entropy characterization has not been previously studied.
- Measurements from a production network show that the entropies of flow elements (external IP address, external port, campus IP address, campus port) and tuples have particular characteristics.
- Findings show that entropies may widely vary in the course of a day. A similar observation applies to the entropy of the campus IP address.
- Building a granular entropy characterization of the individual flow elements can help detect anomalies.
- Data shows that certain attacks produce entropies that deviate from the expected patterns.
- The entropy of the 3-tuple {external IP, campus IP, campus port} is high and consistent over time. A deviation from this pattern is an encouraging anomaly indicator.
- Strong negative and positive correlations exist between some entropy time-series of flow elements.

Methodology

- The border router connects the campus network to the Internet Service Provider (ISP) / Internet.
- The NAT device translates private IP addresses to a single public IP address (campus IP).
- The campus network serves ~2,000 users:
  - ✓ 15 buildings and tens of different departments.
  - ✓ 20 general-purpose computer laboratories and staff offices.



- The metering point is the single border router (Cisco ASR 1000 series) and the traffic direction is inbound.
- Flow information is ready for export when:
  - ✓ Flow is inactive for a certain time (i.e., no new packets received for the flow during the last 15 seconds);
  - ✓ The flow is long lived (active) and its duration is greater than the active timer (1 minute);
  - ✓ TCP flag indicates that the flow is terminated (i.e., FIN, RST flag are received.

Methodology

- For each flow, the router exports:
  - ✓ Source and destination IP addresses;
  - ✓ Source and destination ports;
  - ✓ layer-4 protocol;
  - ✓ TCP flags observed during the connection.
- Connection statistics include number of packets, number of bytes, bytes per packet, and flow duration.
- The information is collected by the flow collector, which implements NetFlow protocol version 9.
- The collector organizes flow data in five-minute time slots; data analysis is conducted for each individual time slot.
- For each external (campus) IP address (port) x_i , the probability $p(x_i)$ is calculated:

$$p(x_i) = \frac{\text{Flows with } x_i \text{ as 3-tuple}}{\text{Total number of flows}}$$

- This work also calculates the entropy of the 3-tuple (external IP, campus IP, campus port). For a given 3-tuple x_i , the corresponding probability is calculated:

$$p(x_i) = \frac{\text{Flows with } x_i \text{ as external (campus) IP addr. (port)}}{\text{Total number of flows}}$$

- For each time slot, the entropy of flow elements is computed:

$$H(X) = \sum_{i=1}^{N} p(x_i)\log_2\left(\frac{1}{p(x_i)}\right)$$

- For each time slot, the five normalized entropies are computed.
- Let $Y_{i,j}$ denote the normalized entropy of distribution I (e.g., campus IP address) observed in time slot j, and Y_i denote the time-series of normalized entropy values for distribution i.
- Given the Y_is, the pairwise correlation coefficients between every pair of time-series vectors Y_i and $Y_{i'}$, are computed:

$$r_{i,i'} = \frac{\sum_j Y_{i,j}Y_{i',j} - n\overline{Y_i}\,\overline{Y_{i'}}}{(n-1)\sigma_{Y_i}\sigma_{Y_{i'}}}$$

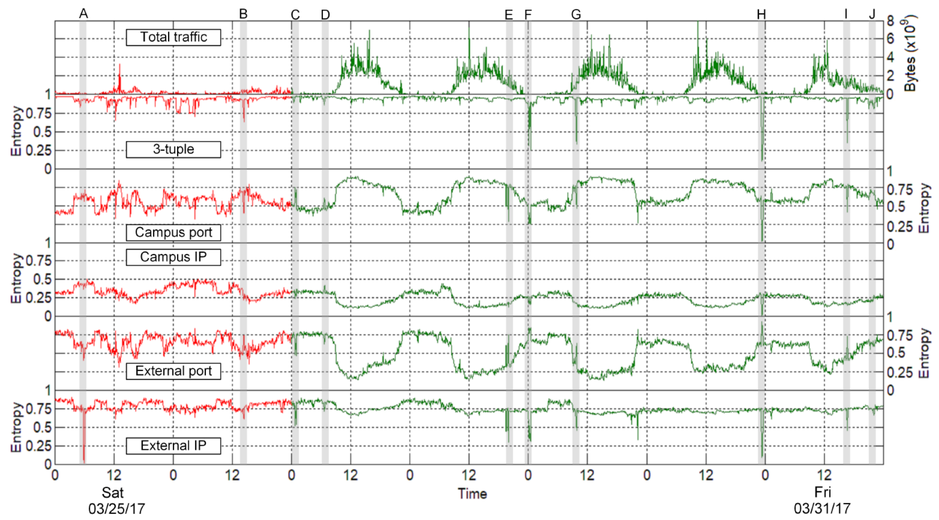
- The rate of change of the entropies is also approximated as an indicator of anomalies.
- A simple approximation of the derivative of Y_i with respect to time is computed as the diff. between consecutive time slots j and j + 1:

$$\Delta Y_{i,j} = Y_{i,j+1} - Y_{i,j}.$$

Acknowledgement

- This work was supported by the National Science Foundation (NSF), Grants 1822567 and 1829698.

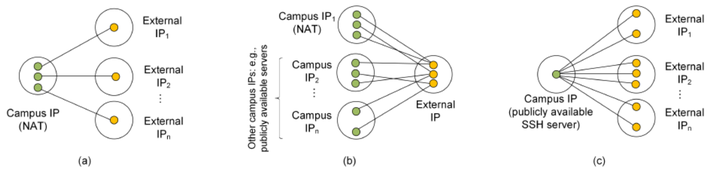
Results



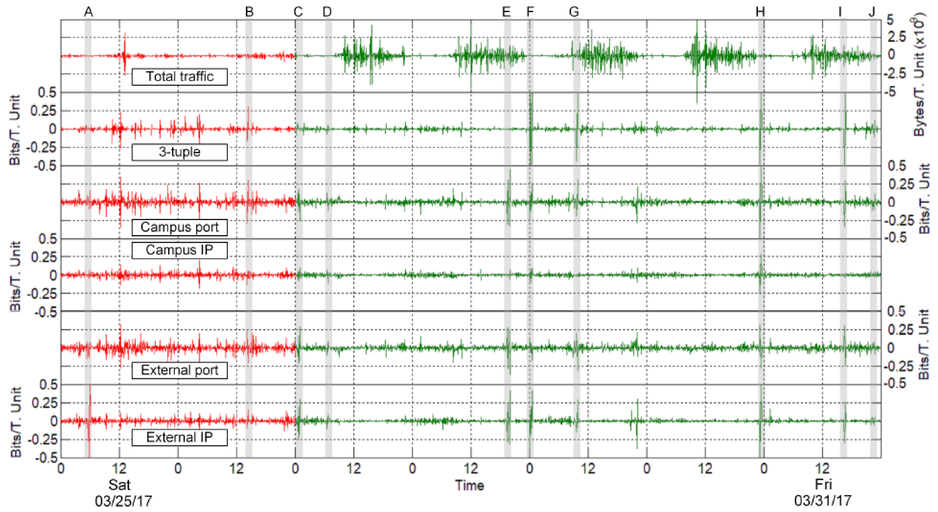
Entropy time-series for the small/medium-sized network. Anomalies are labeled with letters A through J.

CORRELATION OF ENTROPY TIME-SERIES.

	Campus IP	Campus port	External IP	External port	Total traffic
Weekday					
3-tuple	0.23	0.1	0.6	-0.02	-0.05
Campus IP		-0.85	0.6	0.89	-0.8
Campus port			-0.37	-0.98	0.78
External IP				0.45	-0.36
External port					-0.81
Weekend					
3-tuple	-0.23	-0.12	0.56	0.06	-0.03
Campus IP		0.15	-0.38	0.06	-0.38
Campus port			-0.48	-0.93	0.31
External IP				0.48	-0.05
External port					-0.39



(a) Typical flow pattern. (b) Flow pattern showing a unique external IP address generating multiple flows to multiple campus IP addresses. (c) Flow pattern where each of the multiple external IP addresses generates multiple flows to a campus IP address, single campus port (SSH).



Rate of change for the time-series.