Detecting Advanced Network Threats Using a Similarity Search

AIMS 2016

Wednesday 22nd June, 2016

Milan Čermák
Pavel Čeleda

State of the Art of Network Data Analysis

Methods classification based on the detection approach

- Statistical
- Classification based
- Clustering & outlier-based
- Soft computing
- Knowledge-based
- Combination learners

State of the Art of Network Data Analysis

Methods classification based on the detection approach

- Statistical
- Classification based
- Clustering & outlier-based
- Soft computing
- Knowledge-based
- Combination learners

Attack example: *Dictionary attack on the SSH service*

Duration	Protocol	Src IP:Port		Dst IP:Port	Packets	Bytes
1.310	TCP	147.251.AA.BB:49297	->	147.251.CC.DD:22	12	1197
0.269	TCP	147.251.AA.BB:49320	->	147.251.CC.DD:22	11	1157
0.436	TCP	147.251.AA.BB:49329	->	147.251.CC.DD:22	11	1157
0.196	TCP	147.251.AA.BB:49358	->	147.251.CC.DD:22	11	1173
0.155	TCP	147.251.AA.BB:49308	->	147.251.CC.DD:22	11	1157

State of the Art of Network Data Analysis

Methods classification based on the detection approach

- Statistical
- Classification based
- Clustering & outlier-based
- Soft computing
- Knowledge-based
- Combination learners

Attack example: *Dictionary attack on the SSH service*

Duration	Protocol	Src IP:Port		Dst IP:Port	Packets	Bytes
8.157	TCP	147.251.AA.BB:49368	->	147.251.CC.DD:22	142	44441
5.501	TCP	147.251.AA.BB:49379	->	147.251.CC.DD:22	99	30389
14.227	TCP	147.251.AA.BB:49367	->	147.251.CC.DD:22	239	76837
6.722	TCP	147.251.AA.BB:49369	->	147.251.CC.DD:22	119	36981
5.429	TCP	147.251.AA.BB:49372	->	147.251.CC.DD:22	98	29865

State of the Art of Network Data Analysis

Methods classification based on the detection approach

- Statistical
- Classification based
- **Clustering & outlier-based**
- Soft computing
- **Knowledge-based**
- Combination learners

Attack example: *Dictionary attack on the SSH service*

Duration	Protocol	Src IP:Port		Dst IP:Port	Packets	Bytes
8.157	TCP	147.251.AA.BB:49368	->	147.251.CC.DD:22	142	44441
5.501	TCP	147.251.AA.BB:49379	->	147.251.CC.DD:22	99	30389
14.227	TCP	147.251.AA.BB:49367	->	147.251.CC.DD:22	239	76837
6.722	TCP	147.251.AA.BB:49369	->	147.251.CC.DD:22	119	36981
5.429	TCP	147.251.AA.BB:49372	->	147.251.CC.DD:22	98	29865

Similarity Searching

Why?

- Almost every network anomaly detection method utilize some kind of a similarity.
- Possibility of variability in network anomaly characteristics in opposition to the exact match approach.
- Query-by-example principle.



CSIRT-MU

Aim of the Research

Use similarity search techniques for detecting advanced network threats based on similarity of traffic behaviour patterns.

Research Question I.

How can we characterize similarity in network traffic?

Aim of the question

Understanding of network traffic behaviour patterns and their mutual relations from the perspective of a similarity.

Research areas

1. *Definition of behaviour patterns providing reasonable amount of information for a similarity comparison.*
2. *Specification of methods for a similarity comparison of defined behaviour patterns.*

Research Question II.

How can similarity search techniques be utilized for detecting network anomalies?

Aim of the question

Research of transformation possibilities of fundamental methods for network anomalies detection into the similarity search concept.

Research areas

1. *Creation of a collection of behaviour patterns representing selected network anomalies.*
2. *Definition of a technology concept for a network traffic classification using similarity of behaviour patterns.*
3. *Concept evaluation and comparison with other common network threats detection methods.*

Research Question III.

What possibilities do the similarity search techniques have for detecting advanced network threats?

Aim of the question

Utilization of the proposed network anomaly detection approach for detection of advanced network threats.

Research areas

1. *Improvement of the approach by verification of different distance functions and results representation.*
2. *Identification of smaller behaviour patterns and their combinations based on general models of network attacks.*

Proposed Approach – RQ I.

Understanding of network traffic characteristics

- Study of publications focused to network anomalies detection.
- Evaluation of observed characteristics using public datasets and live network analysis.

Specification of behaviour patterns and distance functions

- Utilization of Bro and IP flow monitoring systems.
- Two patterns forms: *aggregated* and *sequential*.
- Utilization of the Metric Similarity Search Implementation Framework (MESSIF).

Proposed Approach – RQ II.

Preparation of annotated behaviour patterns

- Analysis of current network attacks and anomalies observed within live network traffic.

Proof-of-concept framework

- Real-time *kNN*-classification of ongoing traffic.
- Classification based on the similarity with annotated patterns.

Verification of the proposed approach

- Use of simulated network attacks within virtual environment.
- Comparison with common anomaly detection approaches (Snort, Bro, Flowmon ADS, ...).

Proposed Approach – RQ III.

Optimization of similarity search attributes

- Complex study of impacts and the possibilities of similarity search techniques to advanced network threat detection.
- Evaluation of different characteristics of similarity searches and various representations of network behaviour patterns.

Utilization of network security anomaly model

- Based on patterns corresponding to the attack phases instead of the whole attack.
- Utilization of general models of network attacks.

Preliminary Results

Detection of SSH brute-force attacks based on simple similarity of behaviour patterns

- Three patterns based on the average network behaviour of common attack tools (*medusa*, *ncrack*, ...):

	#orig_pkts /#flows	*#resp_pkts /#flows*	*#orig_bytes /#flows*	*#resp_bytes /#flows*	*time /#flows*
medusa vector	15	20	2350	3500	4

- Utilization of simple quadratic form distance function:

$$d_M(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T \cdot M \cdot (\vec{x} - \vec{y})}$$

Preliminary Results

Detection of SSH brute-force attacks based on simple similarity of behaviour patterns

- Three patterns based on the average network behaviour of common attack tools (*medusa*, *ncrack*, ...):

	#orig_pkts /#flows	*#resp_pkts /#flows*	*#orig_bytes /#flows*	*#resp_bytes /#flows*	*time /#flows*
medusa vector	15	20	2350	3500	4

- Utilization of simple quadratic form distance function:

$$d_M(\vec{x},\vec{y}) = \sqrt{(\vec{x}-\vec{y})^T \cdot M \cdot (\vec{x}-\vec{y})}$$

- Ability to identify all variants of attack tools settings.
- Better results than clustering based detection approach.
- Practically no false positives.

Summary and Expected Results

- Proposal of methods for a supervised detection of network anomalies based on the similarity of behaviour patterns.
- Proof-of-concept framework for a network anomalies and attacks detection in a real-time.
- The evaluation of proposed approach and its comparison with other commonly used network traffic anomaly detection methods.
- The description of effects of different similarity search methods to a detection of advanced network threats.

DETECTING ADVANCED NETWORK THREATS USING A SIMILARITY SEARCH

Milan Čermák

cermak@ics.muni.cz