Distributed Anomaly Detection with Network Flow Data
Detecting Network-wide Anomalies

Carlos García C.¹  Andreas Vöst²  Jochen Kögel ²

¹TU Darmstadt
Telecooperation Group & CASED

²IsarNet SWS GmbH

2015-07-24
Table of Contents

1. Securing Complex Networks
2. Discovering Anomalies in Flows
3. Scalable Distributed System
4. Exemplary Results
5. Summary
## Table of Contents

1. Securing Complex Networks
2. Discovering Anomalies in Flows
3. Scalable Distributed System
4. Exemplary Results
5. Summary
The Importance of Network Security

- Computer networks are crucial to daily life
  - banking systems, power plants, your office
- Attacks are more sophisticated and widespread
- How do we protect networks?
- Proactive security is not sufficient (e.g. firewalls)
The Importance of Network Security

- Computer networks are crucial to daily life
  - banking systems, power plants, your office
- Attacks are more sophisticated and widespread
- How do we protect networks?
- Proactive security is not sufficient (e.g. firewalls)
The Importance of Network Security

- Computer networks are crucial to daily life
  - banking systems, power plants, your office
- Attacks are more sophisticated and widespread
- How do we protect networks?
- Proactive security is not sufficient (e.g. firewalls)
The Importance of Network Security

- Computer networks are crucial to daily life
  - banking systems, power plants, your office
- Attacks are more sophisticated and widespread
- How do we protect networks?
- Proactive security is not sufficient (e.g. firewalls)
The Importance of Network Security

- Computer networks are crucial to daily life
  - banking systems, power plants, your office
- Attacks are more sophisticated and widespread
- How do we protect networks?
- Proactive security is not sufficient (e.g. firewalls)
The Importance of Network Security

- Computer networks are crucial to daily life
  - banking systems, power plants, your office
- Attacks are more sophisticated and widespread
- How do we protect networks?
- Proactive security is not sufficient (e.g. firewalls)
The Importance of Network Security

- Computer networks are crucial to daily life
  - banking systems, power plants, your office
- Attacks are more sophisticated and widespread
- How do we protect networks?
- Proactive security is not sufficient (e.g. firewalls)
The Importance of Network Security

- Computer networks are crucial to daily life
  - banking systems, power plants, your office
- Attacks are more sophisticated and widespread
- How do we protect networks?
- Proactive security is not sufficient (e.g. firewalls)
Reactive Security

- Security cannot be guaranteed
- Detect security and policy violations after their occurrence

Scenario: Small Network
Reactive Security

- Security cannot be guaranteed
- Detect security and policy violations after their occurrence

Scenario: Small Network
Reactive Security

- Security cannot be guaranteed
- Detect security and policy violations after their occurrence

Scenario: Small Network
Reactive Security

- Security cannot be guaranteed
- Detect security and policy violations after their occurrence
Reactive Security

- Security cannot be guaranteed
- Detect security and policy violations after their occurrence

Scenario: Small Network

- One common point of ingress
- Complete view of the network
- Flows captured in one place
Reactive Security

**Scenario: Large Network**

- A distributed monitoring system is required
- Reactive security utilizing **distributed IDSs**
Reactive Security

Scenario: Large Network

- A distributed monitoring system is required
- Reactive security utilizing distributed IDSs
Reactive Security

Scenario: Large Network

- A distributed monitoring system is required
- Reactive security utilizing distributed IDSs
Reactive Security

A distributed monitoring system is required
Reactive security utilizing distributed IDSs
Reactive Security

**Scenario: Large Network**

- Multiple ingress points
- Partial view of the network
- Flows aggregated in many places

A distributed monitoring system is required

Reactive security utilizing **distributed IDSs**
Reactive Security

Scenario: Large Network

- Multiple ingress points
- Partial view of the network
- Flows aggregated in many places

A distributed monitoring system is required
Reactive security utilizing distributed IDSs
Distributed Intrusion Detection

Flow Monitoring
- Distributed monitoring with IsarFlow
- To collect, aggregate and perform anomaly detection

Anomaly Detection
- To detect unknown problems
  - Attacks or intrusions
  - Irregular operation
- Detect anomalies present in flows
Distributed Intrusion Detection

Flow Monitoring

- Distributed monitoring with IsarFlow
- To collect, aggregate and perform anomaly detection

Anomaly Detection

- To detect unknown problems
  - Attacks or intrusions
  - Irregular operation
- Detect anomalies present in flows
Distributed Intrusion Detection

Flow Monitoring
- Distributed monitoring with IsarFlow
- To collect, aggregate and perform anomaly detection

Anomaly Detection
- To detect unknown problems
  - Attacks or intrusions
  - Irregular operation
- Detect anomalies present in flows
Distributed Intrusion Detection

Flow Monitoring
- Distributed monitoring with IsarFlow
- To collect, aggregate and perform anomaly detection

Anomaly Detection
- To detect unknown problems
  - Attacks or intrusions
  - Irregular operation
- Detect anomalies present in flows
# Table of Contents

1. Securing Complex Networks
2. Discovering Anomalies in Flows
3. Scalable Distributed System
4. Exemplary Results
5. Summary
Anomalies in Network Flows

Flow Anomaly

Any network traffic exhibiting unexpected or undesired patterns of communication in flows.

Common Network Anomalies

- Malicious Activity
  - (D)DoS
  - Port Scans
  - Worms & Botnets
- Operational Problems
  - Alpha Flows
  - Ingress Shifts (Outages)
  - Large quantities of small packets
- Noteworthy Events
  - Flash Crowds
  - Bittorrent Traffic
Anomalies in Network Flows

Flow Anomaly
Any network traffic exhibiting unexpected or undesired patterns of communication in flows.

Common Network Anomalies
- Malicious Activity
  1. (D)DoS
  2. Port Scans
  3. Worms & Botnets
- Operational Problems
  1. Alpha Flows
  2. Ingress Shifts (Outages)
  3. Large quantities of small packets
- Noteworthy Events
  1. Flash Crowds
  2. Bittorrent Traffic
Anomalies in Network Flows

Flow Anomaly

Any network traffic exhibiting unexpected or undesired patterns of communication in flows.

Common Network Anomalies

- Malicious Activity
  1. (D)DoS
  2. Port Scans
  3. Worms & Botnets

- Operational Problems
  1. Alpha Flows
  2. Ingress Shifts (Outages)
  3. Large quantities of small packets

- Noteworthy Events
  1. Flash Crowds
  2. Bittorrent Traffic
Categories of Anomalies

## Anomalies in Network Flows

### Flow Anomaly

Any network traffic exhibiting unexpected or undesired patterns of communication in flows.

### Common Network Anomalies

- **Malicious Activity**
  1. (D)DoS
  2. Port Scans
  3. Worms & Botnets

- **Operational Problems**
  1. Alpha Flows
  2. Ingress Shifts (Outages)
  3. Large quantities of small packets

- **Noteworthy Events**
  1. Flash Crowds
  2. Bittorrent Traffic
The Nature of Network Flows

- Highly dimensional data
- Data can be both numerical and categorical (e.g., protocol names)
- Do not contain network payload
- Often contain sampled data
- Vast quantities of information

Intrusion detection is difficult in this problem space
Feature extraction and summarization is required

Feature Extraction Strategies

- Volume-based feature extraction
- Entropy-based feature extraction
The Nature of Network Flows

- Highly dimensional data
- Data can be both numerical and categorical (e.g., protocol names)
- Do not contain network payload
- Often contain sampled data
- Vast quantities of information

- Intrusion detection is difficult in this problem space
- Feature extraction and summarization is required

Feature Extraction Strategies

- Volume-based feature extraction
- Entropy-based feature extraction
The Nature of Network Flows

- Highly dimensional data
- Data can be both numerical and categorical (e.g., protocol names)
- Do not contain network payload
- Often contain sampled data
- Vast quantities of information

- Intrusion detection is difficult in this problem space
- Feature extraction and summarization is required

Feature Extraction Strategies

- Volume-based feature extraction
- Entropy-based feature extraction
The Nature of Network Flows

- Highly dimensional data
- Data can be both numerical and categorical (e.g., protocol names)
- Do not contain network payload
- Often contain sampled data
- Vast quantities of information

- Intrusion detection is difficult in this problem space
- Feature extraction and summarization is required

Feature Extraction Strategies

- Volume-based feature extraction
- Entropy-based feature extraction
Entropy-based Feature Analysis

Why is Entropy Interesting?
- Every flow feature can be summarized with its entropy
  - e.g., source and destination IP, source and destination port
- Compact representation of all features

Entropy ($H$):
- Degree of randomness
- Maximum if all values are equal
- Minimal if probability mass concentrates on one value

Shannon Entropy ($H$)

\[ X = \{ n_i, i = 1, \ldots, N \} \]
\[ H(X) = - \sum_{i=1}^{N} \left( \frac{n_i}{N} \right) \log_2 \left( \frac{n_i}{N} \right) \]
\[ 0 < H(X) < \log_2 N \]
Entropy-based Feature Analysis

Key Property of Entropy

- Entropy measures the concentration or dispersal of a distribution

Normal Traffic

Port Scan Traffic

Entropy measures the concentration or dispersal of a distribution.
Entropy-based Feature Analysis

Key Property of Entropy
- Entropy measures the concentration or dispersal of a distribution

\[ H(\text{Normal Traffic}) > H(\text{Port Scan Traffic}) \]
Entropy Time Series

Anomaly Detection using Entropy

1. Select a time window
2. For each window:
   - Build histograms of the desired features
   - Calculate the Entropy of each histogram
   - Build a time series of the entropies
3. Choose algorithm to detect unusual patterns
   - K-Means clustering
   - Gaussian Mixture Models (GMMs)
   - Subspace Method
Entropy Time Series

Anomaly Detection using Entropy

1. Select a **time window**
2. For each window:
   - Build histograms of the desired features
   - Calculate the Entropy of each histogram
   - Build a time series of the entropies
3. Choose algorithm to detect unusual patterns
   - K-Means clustering
   - Gaussian Mixture Models (GMMs)
   - Subspace Method
Entropy Time Series

Anomaly Detection using Entropy

1. Select a **time window**
2. For each window:
   1. Build histograms of the desired features
   2. Calculate the **Entropy** of each histogram
   3. Build a **time series** of the entropies
3. Choose algorithm to detect unusual patterns
   1. K-Means clustering
   2. Gaussian Mixture Models (GMMs)
   3. Subspace Method
Entropy Time Series

Anomaly Detection using Entropy

1. Select a **time window**
2. For each window:
   1. **Build histograms** of the desired features
   2. **Calculate** the **Entropy** of each histogram
   3. **Build a time series** of the entropies
3. Choose algorithm to detect unusual patterns
   1. K-Means clustering
   2. Gaussian Mixture Models (GMMs)
   3. Subspace Method
Entropy Time Series

**Anomaly Detection using Entropy**

1. Select a **time window**
2. For each window:
   1. **Build histograms** of the desired features
   2. **Calculate the Entropy** of each histogram
   3. **Build a time series** of the entropies
3. Choose algorithm to detect unusual patterns
   1. K-Means clustering
   2. Gaussian Mixture Models (GMMs)
   3. Subspace Method
Entropy Time Series

Anomaly Detection using Entropy

1. Select a **time window**
2. For each window:
   1. **Build histograms** of the desired features
   2. **Calculate the Entropy** of each histogram
   3. **Build a time series** of the entropies
3. Choose algorithm to detect unusual patterns
   1. K-Means clustering
   2. Gaussian Mixture Models (GMMs)
   3. Subspace Method
Table of Contents

1. Securing Complex Networks
2. Discovering Anomalies in Flows
3. Scalable Distributed System
4. Exemplary Results
5. Summary
Distributed Monitoring System

Exemplary architecture: The **IsarFlow** Network Monitoring System

- **Distributed** collection, storage and data analysis
  - Scales very well with more analyzers
  - No need to send flow data across WAN
- Detection Algorithms must also scale in a distributed way
Combination of Models

How to derive models of normality in a distributed system?

Model Merging

Model Composition
Combination of Models

**Model Merging**
- Calculate features locally
- Exchange features with other analyzers
- Determine global model of normality - based on all feature information

**Model Composition**
- Calculate features locally
- Train classifier with local features
- Classify traffic with local classifier
- Forward local classification result to evaluation instance (Composer)
Combination of Models

### Model Merging

- **Global Model**
- **All analyzer utilize same detection model**
- **Learned model can be exchanged**
  - Necessity to exchange feature information
  - Features need to be interchangeable

### Model Composition

- **Local model might be more precise**
- **No feature exchange necessary**
- **Smaller overhead**
  - Model might not be interchanged
  - Composer has to be trained
Table of Contents

1. Securing Complex Networks
2. Discovering Anomalies in Flows
3. Scalable Distributed System
4. Exemplary Results
5. Summary
Example: PortScan Entropy Fingerprint
Example: PortScan Entropy Fingerprint
Table of Contents

1. Securing Complex Networks
2. Discovering Anomalies in Flows
3. Scalable Distributed System
4. Exemplary Results
5. Summary
Summary and Outlook

Summary

- Reactive traffic monitoring is crucial
- Challenges in large enterprise networks
  - Large amount of unsampled flow data
  - Needs distributed collection and data processing
- Entropy as promising feature
  - Difficult to cope with distributed data
  - Approach requires efficient data combination

Outlook

- Thorough study of flow data from a large enterprise network
- Evaluation of feature extraction and classifiers
- Study of detection precision and accuracy
THANK YOU FOR YOUR ATTENTION
Example: DDoS Reflector Attack detection
Example: DDoS Reflector Attack detection
Example: DDoS Reflector Attack detection
Example: DDoS Reflector Attack detection
Example: DDoS Reflector Attack detection

![Graph showing H(DST Port) over time with marked anomalies at certain time points.](image-url)
Example: DDoS Reflector Attack detection

![Graph showing H(PacketCount) over time with peaks at 10:00 and 11:06]

**H(PacketCount)**

- **Time:** 08:53, 09:26, 10:00, 10:33, 11:06
- **Y-axis:** H(PacketCount) from 0.0 to 4.5
- **Legend:** All Flows (red)
Example: DDoS Reflector Attack detection
Example: Worm Scan detection
Example: Worm Scan detection
Example: Worm Scan detection
Example: Worm Scan detection
Example: Worm Scan detection
Example: Worm Scan detection
Example: Worm Scan detection