Network Management with Data Analytics

Lautaro Dolberg
Supervisor: Prof. Dr. Thomas Engel

UNIVERSITÉ DU LUXEMBOURG

SnT

Outline

Outline

1. Introduction
2. Multidimensional Aggregation Monitoring
3. Case Study: DNS Monitoring
4. Application Awareness
5. Conclusion

Network Management

Definition

In computer networks, network management is the operation, administration, maintenance, and provisioning (OAMP) of networked systems[a]

[a] Source: Cisco Systems

FCAPS

A common way of characterizing network management functions is FCAPS. **Fault, Configuration, Accounting, Performance and Security.**

Fault Management

Fault Management is established by monitoring different things for abnormal behaviour.

Data Analytics

Data Analytics

"Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate..." [a]

---

[a] John Tukey,
The Future of Data Analysis, 1961

Data Analytics

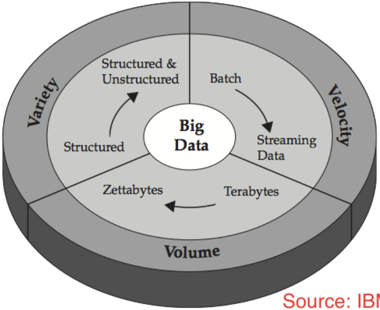
Data Analytics

"Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate..." [a]

[a] John Tukey, The Future of Data Analysis, 1961

Big Data

Volume, Variety, and Velocity



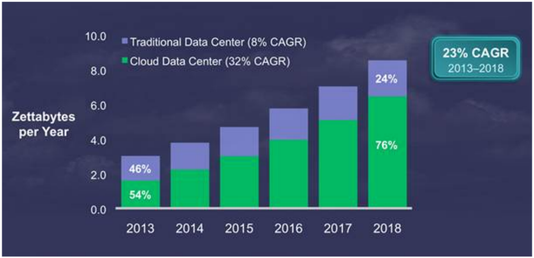
Source: IBM

Big Data Analytics

Combines data analytics to the challenges of Big Data

Traffic Volume Growth

Data Centres Traffic



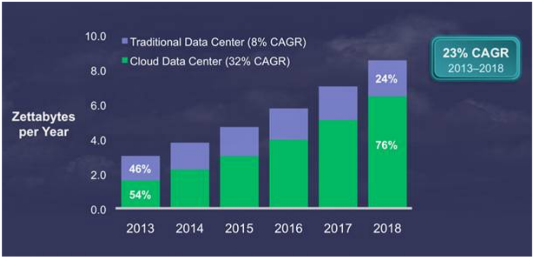
Source: Cisco Global Cloud Index, 2013–2018

Generalized Traffic

The data centre traffic has dramatically grown in the last years.

Traffic Volume Growth

Data Centres Traffic



Source: Cisco Global Cloud Index, 2013–2018

Cloud-based Traffic

In particular data intensive applications require more bandwidth than only CPU intensive

Network Management Challenges

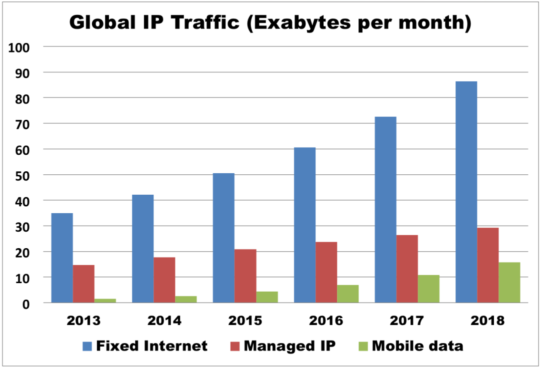
Support for Data Analytics

- Resource Allocation (Virtualized Environments)
- Network Bandwidth (Data intensive)
- Flexibility

Cloud Environments

In particular support for data intensive applications in the cloud

Internet Traffic Volume Growth



Data Analytics for Security

The scale and complexity of the threats require strong techniques for efficient monitoring.

Security Perspective

During year 2014 DDoS Attacks:

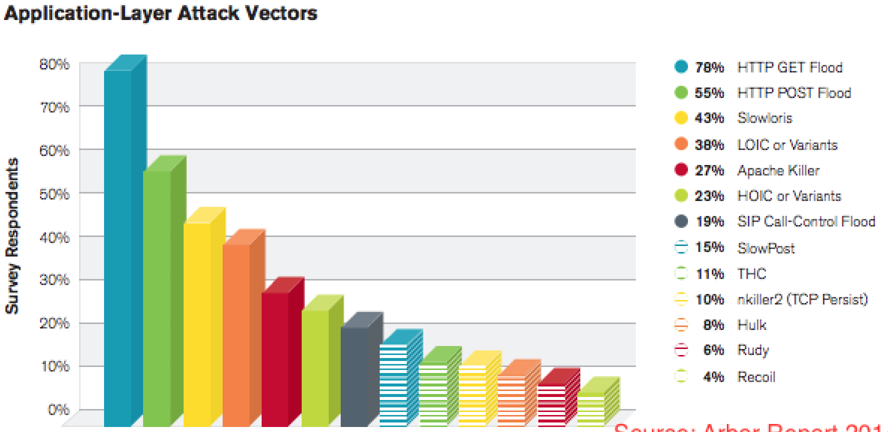
- Remain the number one operational threat seen by respondents (Arbor 2014).
- Multiple respondents report very large attacks above the 100Gbps threshold (Kaspersky Labs).

Increase in complexity of attacks

- Advanced persistent threats are increasingly common (Arbor 2014)

Diversity of Attack Vectors

Attacks Vector Reported in 2014



Source: Arbor Report 2014

Network Security

Network Security Monitoring

- Collect network-derived data
- Analyze to find anomalies
- Evaluate anomalies and *find out attacks*

Network Monitoring is essential for detecting attacks

Challenges

- No Silver Bullet $\leftarrow$ Attack diversity
- Volume of information can be overwhelming
- Attacks are becoming more complex

Research Questions: Network Security

Network Monitoring

Is an essential part in network security for detection of attacks

In the context of network monitoring

- How diverse information from network sources can be:
  - extracted
  - stored
  - aggregated to outline significant events
- **Detection of anomalies**

Research Questions: Network Security

Network Monitoring

Is an essential part in network security for detection of attacks

In the context of network monitoring

- How diverse information from network sources can be:
  - extracted
  - stored
  - aggregated to outline significant events
- **Detection of anomalies**

Example

A NetFlow log has at least 12 fields per record

Research Questions: Case Studies

Context: Anomaly Detection

- TCP/IP Networks:
  - Anomalies in IP Traffic Patterns

Research Questions: Case Studies

Context: Anomaly Detection

- TCP/IP Networks:
  - Anomalies in IP Traffic Patterns
- Domain Name System (DNS) :
  - Find out anomalies in the association between names and IP addresses

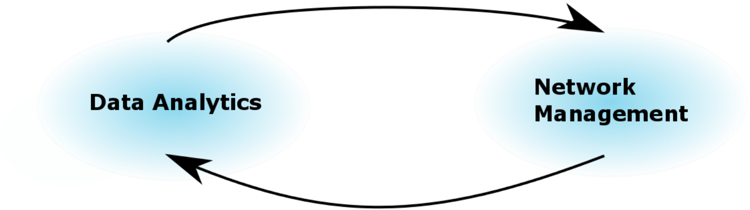
Research Questions: Case Studies

Context: Anomaly Detection

- TCP/IP Networks:
  - Anomalies in IP Traffic Patterns
- Domain Name System (DNS) :
  - Find out anomalies in the association between names and IP addresses
- Crowd-sourced Position-based Applications (Routing)
  - Study the dynamics of the reported positions of moving vehicles

Research Questions: Network Management

In the context of Network Management



Research Questions: Network Management

In the context of Network Management

- How *application awareness* can be brought to a network level for:
  - Traffic Analysis
    - Application
    - Instance
    - User/Host

Research Questions: Network Management

In the context of Network Management

- How *application awareness* can be brought to a network level for:
  - Traffic Analysis
    - Application
    - Instance
    - User/Host
  - Network Management
    - Data Intensive
    - Resource allocation
    - Enrich QoS (Quality of Service) rules
    - Security

Research Questions: Network Management

In the context of Network Management

- How *application awareness* can be brought to a network level for:
  - Traffic Analysis
    - Application
    - Instance
    - User/Host
  - Network Management
    - Data Intensive
    - Resource allocation
    - Enrich QoS (Quality of Service) rules
    - Security
- **Leverage network management with application awareness**

Outline

SnT

Network Monitoring

What if?

- Traces from TCP/IP communications
- What do we see if we group per sub network? (activity)
- What ports are the most used? (applications)
- Is there any significant changes in terms of usage?

Challenges

- Human expertise
- Time consuming
- Costly process
- Scalability

SnT

Multidimensional Monitoring

How do we handle multidimensional data from network events?

SnT

Multidimensional Monitoring

How do we handle multidimensional data from network events?

- Preserve the hierarchy of Data (DNS, IP, Coordinates, SIP, HTTP, ...).
- Flexible Granularity
- Minimize information loss due to aggregation
- Reduce the scale

Multidimensional Monitoring

How do we handle multidimensional data from network events?

- Preserve the hierarchy of Data (DNS, IP, Coordinates, SIP, HTTP, . . .).
- Flexible Granularity
- Minimize information loss due to aggregation
- Reduce the scale

Aggregation

- Scalable way to outline relevant facts

SnT

Multidimensional Monitoring

How do we handle multidimensional data from network events?

- Preserve the hierarchy of Data (DNS, IP, Coordinates, SIP, HTTP, ...).
- Flexible Granularity
- Minimize information loss due to aggregation
- Reduce the scale

Aggregation

- Scalable way to outline relevant facts
- Custom Units (e.g. traffic packets, traffic units, records)

SnT

Multidimensional Monitoring

How do we handle multidimensional data from network events?

- Preserve the hierarchy of Data (DNS, IP, Coordinates, SIP, HTTP, ...).
- Flexible Granularity
- Minimize information loss due to aggregation
- Reduce the scale

Aggregation

- Scalable way to outline relevant facts
- Custom Units (e.g. traffic packets, traffic units, records)
- Temporal: time windows split (β)
- Spatial: keep group of events with activity $> \alpha$ *e.g. traffic volume*

State of the Art

Space Partitioning: static vs dynamic

- Static Partitioning: defining partitions that remain unchangeable
  For example a network prefix (/8,/16,/24)
- Dynamic Partitioning: space is partitioned according to given data
  Natural relationships are lost sometimes

Partitioning

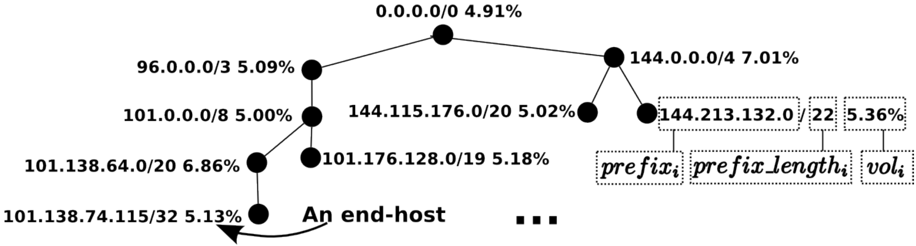
Static: Quadtrees ; Dynamic: Hierarchical Clustering

Data Structure

Approaches such as Relational Data Bases require well structured data and a data model

Simple Aggregation

Aguri Tree Example



- Node Volume
- IP Address & Network Address

K. Cho at Al, "Aguri: An aggregation-based traffic profiler," in *Quality of Future Internet Services*. Springer, 2001.

Outline

Multidimensional Data

Example: NetFlow Capture

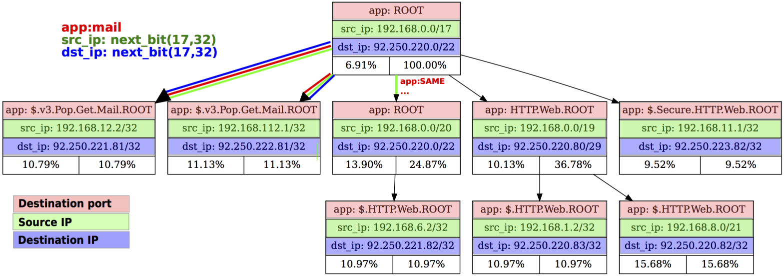
PORT	PROTO	KB	TIME	SOURCE	DEST
80	TCP	1491	2010-02-24 02:20:15	192.168.6.2	92.250.221.82
110	TCP	988	2010-02-24 02:20:19	192.168.8.2	92.250.223.87
443	TCP	902	2010-02-24 02:20:27	192.168.11.2	92.250.220.82
110	TCP	1513	2010-02-24 02:20:29	192.168.112.1	92.250.222.81
80	TCP	1205	2010-02-24 02:20:29	192.168.11.1	92.250.220.82
80	TCP	1491	2010-02-24 02:20:31	192.168.1.2	92.250.220.83
110	TCP	1467	2010-02-24 02:20:39	192.168.12.2	92.250.221.81
80	TCP	927	2010-02-24 02:20:39	192.168.12.2	92.250.220.82
443	TCP	1294	2010-02-24 02:20:39	192.168.11.1	92.250.223.82
110	TCP	940	2010-02-24 02:20:49	192.168.21.2	92.250.221.81
80	TCP	917	2010-02-24 02:20:49	192.168.23.1	92.250.220.82
443	TCP	460	2010-02-24 02:20:59	192.168.26.2	92.250.220.85

Dimensions

- Source IP Address & Network Address
- Destination IP Address & Network Address
- Application

Multidimensional Aggregated Tree

Previous NetFlow Capture Tree Representation



Tree Construction

- Aggregation α: 9%
- Directions to navigate the tree

Tree Construction

Structure

Root node and multiple children, as many as each dimension admit.

Tree navigation

- How to find the right path to insert a node within a tree?
- Direction function indicates the next step
  - Most specific common ancestor between two nodes
  - Longest common prefix match (IP, DNS and other address spaces)
- For IPv4 is a binary function (0,1) as next bit value
- For other features, natural hierarchical classification

Outline

1. Introduction
2. Multidimensional Aggregation Monitoring
   - Multidimensional Trees
   - Evaluation
   - Summary
3. Case Study: DNS Monitoring
4. Application Awareness
5. Conclusion

Data Sets

NetFlow: ISP from Luxembourg (3-D)

IP Addresses	Ports	Duration	Flows
279815	64470	26 days	60,000/sec

GPS: Luxembourg Urban Scenario Simulation (2-D)

Total Cars	Mean Speed	Mean Car Trip Time	Reporting Interval
138260	11.17	11 min	1 second

DNS: Passive DNS DB (2-D)

	Domains	IP Address
Normal	661968	164559
Malicious	173066	174619
Total	835034	339178

Scalability

Aggregation Results

	Average tree size before aggregation	Average tree size after aggregation
Netflow	3288	90
DNS	8600	53
SIP	1077	18
Geographical	14664	45

Table 1: Tree size reduction using aggregation (average on all time windows) with $\alpha = 0.05$

Scalability

Aggregation Results

	Average tree size before aggregation	Average tree size after aggregation
Netflow	3288	90
DNS	8600	53
SIP	1077	18
Geographical	14664	45

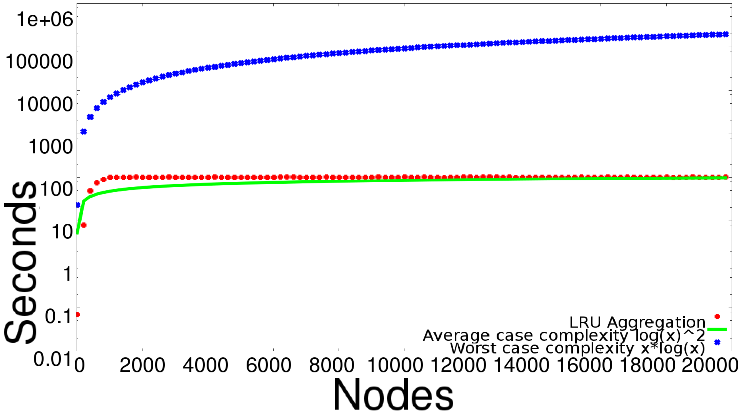
Table 1: Tree size reduction using aggregation (average on all time windows) with $\alpha = 0.05$

Scalability

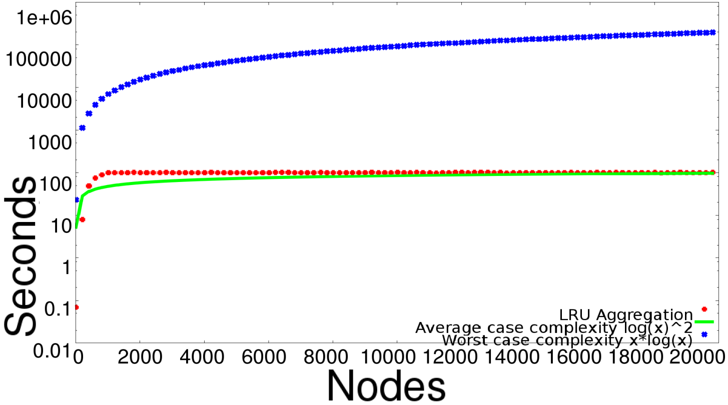
- Aggregation reduced the scale of data in at least one order of magnitude
- An ISP generates 60,000 records per second
- With aggregation this is collapsed to approximately 1800 nodes

Performance

Running Time



Running Time



Aggregation running times $O(n \times \log(n))$ are far from the worst case complexity calculations $O(n^2)$

Outline

1. Introduction
2. Multidimensional Aggregation Monitoring
   - Multidimensional Trees
   - Evaluation
   - Summary
3. Case Study: DNS Monitoring
4. Application Awareness
5. Conclusion

Conclusions

Summary

- Validated aggregation of heterogeneous data (Mutidimensional)
- Data Driven Granularity
- Scalable Aggregation
- https://github.com/ldolberg/mam
- Extensible to other data sources and dimensions

Minimize Data Loss

In the next section...

Outline

DNS traffic reflects Internet activities and behaviours

- Internet Threats Growing: Phishing, Malware, Spoofed Domains.
- Identify malware behaviour by assessing association time between names and networks.
- As both DNS and IP space follow hierarchical organization.

Questions...

- What can be learned from DNS records related to Internet activities?

SnT

Introduction

DNS traffic reflects Internet activities and behaviours

- Internet Threats Growing: Phishing, Malware, Spoofed Domains.
- Identify malware behaviour by assessing association time between names and networks.
- As both DNS and IP space follow hierarchical organization.

Questions...

- What can be learned from DNS records related to Internet activities?
- Can malicious domains be identified by looking at DNS Records ?

SnT

Introduction

DNS traffic reflects Internet activities and behaviours

- Internet Threats Growing: Phishing, Malware, Spoofed Domains.
- Identify malware behaviour by assessing association time between names and networks.
- As both DNS and IP space follow hierarchical organization.

Questions...

- What can be learned from DNS records related to Internet activities?
- Can malicious domains be identified by looking at DNS Records ?

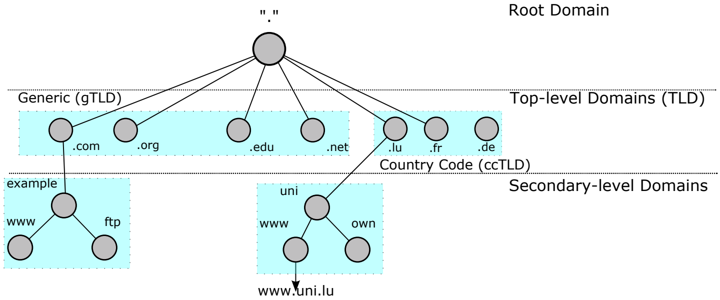
Outline

DNS Background

DNS is an essential service for Internet

- Emerged in 1987 (RFC 1035, 1123, 2181)
- Domain names are labels separated with dots.

Domain Name Hierarchy Example



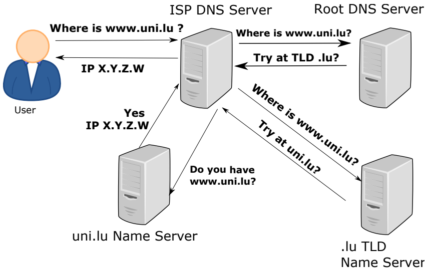
Max depth 127. Limited to 253 characters, label limit 63 characters.

DNS Structure & Procedure

Resolving www.uni.lu

DNS Structure & Procedure

Resolving www.uni.lu



DNS Registers

- The duration of a register is given by its TTL (Time To Live)
- A malicious domains can set a small TTL to avoid blacklisting
- How we can observe changes from a global perspective? (network & domain)

State of the art

Active

- Black Listing: Users report malicious sites and lists are confectioned
- Brute Forcing: Generation of domains

Passive

Feature Extraction: Looks for anomalies in the DNS Responses such as TTL, Record Type, Network Type, etc. (DNSSM, Exposure)

Sources

- L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi, "Exposure: Finding malicious domains using passive dns analysis" in *Network and Distributed System Security Symposium - NDSS*, 2011.
- S. Marchal, J. François, C. Wagner, R. State, A. Dulaunoy, T. Engel, and O. Festor, "DNSSM: A large-scale Passive DNS Security Monitoring Framework", in *IEEE/IFIP Network Operations and Management Symposium*, 2012.

Outline

1. Introduction
2. Multidimensional Aggregation Monitoring
3. Case Study: DNS Monitoring
   - DNS Background
   - Similarity Metric
   - Evaluation
4. Application Awareness
5. Conclusion

MAM for DNS

With MAM is possible to generate aggregated trees combining multiple data

- Two dimensions

MAM for DNS

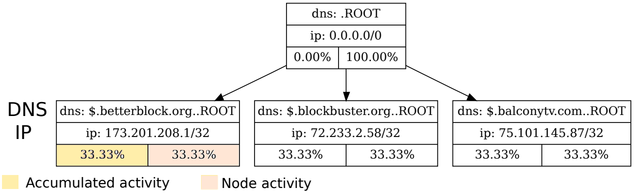
With MAM is possible to generate aggregated trees combining multiple data

- Two dimensions
- Derived from the hierarchically data model (IPV4 & DNS Data Space)

MAM for DNS

With MAM is possible to generate aggregated trees combining multiple data

- Two dimensions
- Derived from the hierarchically data model (IPV4 & DNS Data Space)
- Example



Formal Definition

Assuming a sequence of K Trees

- $S = [T_1, \ldots, T_K]$ representing DNS record association over time split in K

Formal Definition

Assuming a sequence of K Trees

- $S = [T_1, \ldots, T_K]$ representing DNS record association over time split in K
- $n \in T_i$

Formal Definition

Assuming a sequence of K Trees

- $S = [T_1, \ldots, T_K]$ representing DNS record association over time split in K
- $n \in T_i$
  - n^{dns} represents the DNS name

Formal Definition

Assuming a sequence of K Trees

- $S = [T_1, \ldots, T_K]$ representing DNS record association over time split in K
- $n \in T_i$
  - n^{dns} represents the DNS name
  - $n^{\text{ip}} = \left\langle n^{\text{address}}, n^{\text{prefix\_length}} \right\rangle$ represent the IPv4 address as a tuple

Formal Definition

Assuming a sequence of K Trees

- $S = [T_1, \ldots, T_K]$ representing DNS record association over time split in K
- $n \in T_i$
  - n^{dns} represents the DNS name
  - $n^{\text{ip}} = \left\langle n^{\text{address}}, n^{\text{prefix\_length}} \right\rangle$ represent the IPv4 address as a tuple
  - n^{accum} is the accumulated value representing the number of A Records.

Formal Definition

Assuming a sequence of K Trees

- $S = [T_1, \ldots, T_K]$ representing DNS record association over time split in K
- $n \in T_i$
  - n^{dns} represents the DNS name
  - $n^{\text{ip}} = \left\langle n^{\text{address}}, n^{\text{prefix\_length}} \right\rangle$ represent the IPv4 address as a tuple
  - n^{accum} is the accumulated value representing the number of A Records.
- The trees from S are aggregated according to a given α

Formal Definition

Assuming a sequence of K Trees

- $S = [T_1, \ldots, T_K]$ representing DNS record association over time split in K
- $n \in T_i$
  - n^{dns} represents the DNS name
  - $n^{\text{ip}} = \left\langle n^{\text{address}}, n^{\text{prefix\_length}} \right\rangle$ represent the IPv4 address as a tuple
  - n^{accum} is the accumulated value representing the number of A Records.
- The trees from S are aggregated according to a given α

How the similarity between two nodes can be defined?

Similarity

Similarity is positive if: Perfect Match

- $n_1^{dns} \subset n_2^{dns}$ AND $n_1^{ip} \subset n_2^{ip}$

Example

Domain Name	IP
www.uni.lu	192.1.2.1/24
www.uni.lu	192.1.2.2/24

Domain Name	IP
uni.lu	192.168.2.0/24
ftp.uni.lu	192.168.2.1/24

Similarity

Similarity is positive if: Partial Match

- $n_1^{dns} \subset n_2^{dns}$ OR $n_1^{ip} \subset n_2^{ip}$
- $n_2^{dns} \subset n_1^{dns}$ OR $n_2^{ip} \subset n_1^{ip}$

Examples

Domain Name	IP
www.uni.lu	10.1.0.0/8
ftp.uni.lu	10.1.0.0/8

Domain Name	IP
uni.lu	11.1.0.0/8
ftp.uni.lu	10.2.0.0/8

Similarity

Similarity is 0 if: No Match

Both dns and ip are unrelated

Example

Domain Name	IP
www.uni.lu	10.1.0.0/8
www.uni.de	15.1.0.0/5

Similarity Function

DNS Similarity

$$s^{\text{dns}}(n_1, n_2) = \frac{|n_1^{\text{dns}} \cap n_2^{\text{dns}}|}{|n_1^{\text{dns}} \cup n_2^{\text{dns}}|}$$

Example

$$\frac{\{\text{www,uni,lu}\} \cap \{\text{uni,lu}\}}{\{\text{www,uni,lu}\} \cup \{\text{uni,lu}\}}$$

Similarity Function

DNS Similarity

$$s^{\text{dns}}(n_1, n_2) = \frac{|n_1^{\text{dns}} \cap n_2^{\text{dns}}|}{|n_1^{\text{dns}} \cup n_2^{\text{dns}}|}$$

IP Similarity

$$s^{\text{ip}}(n_1, n_2) = 1 - \frac{|n_1^{\text{prefix\_len}} - n_2^{\text{prefix\_len}}|}{32}$$

Example

$$10.1.10.0/8 \; ; \; 10.1.11.0/16 \rightarrow \frac{|8 - 16|}{32}$$

Similarity Function

DNS Similarity

$$s^{\text{dns}}(n_1, n_2) = \frac{|n_1^{\text{dns}} \cap n_2^{\text{dns}}|}{|n_1^{\text{dns}} \cup n_2^{\text{dns}}|}$$

IP Similarity

$$s^{\text{ip}}(n_1, n_2) = 1 - \frac{|n_1^{\text{prefix\_len}} - n_2^{\text{prefix\_len}}|}{32}$$

Volume Similarity

$$s^{\text{vol}}(n_1, n_2) = 1 - 0.01 \times |n_1^{\text{acc\_vol}} - n_2^{\text{acc\_vol}}|$$

Example

$n_1^{\text{acc\_vol}}$, $n_2^{\text{acc\_vol}}$ are percentages

Similarity Function

DNS Similarity

$$s^{\text{dns}}(n_1, n_2) = \frac{|n_1^{\text{dns}} \cap n_2^{\text{dns}}|}{|n_1^{\text{dns}} \cup n_2^{\text{dns}}|}$$

IP Similarity

$$s^{\text{ip}}(n_1, n_2) = 1 - \frac{|n_1^{\text{prefix\_len}} - n_2^{\text{prefix\_len}}|}{32}$$

Volume Similarity

$$s^{\text{vol}}(n_1, n_2) = 1 - 0.01 \times |n_1^{\text{acc\_vol}} - n_2^{\text{acc\_vol}}|$$

Formula

$$s(n_1, n_2) = w_1 \times s^{\text{ip}}(n_1, n_2) + w_2 \times s^{\text{dns}}(n_1, n_2) + w_3 \times s^{\text{vol}}(n_1, n_2)$$
$$w_i \in [0, 1], i \in [1, 2, 3]$$

Smoothing

Smoothing helps considering early time windows

- $n \in T_1, \; m \in T_2$
- We compute m as sim(n,m) is maximum
- $\text{stead}(n) = \text{sim}(n, m) + \mu \times \text{stead}(m)$. With T_0 as base case and $\mu \in \Re$.

Smoothing helps considering early time windows

- $n \in T_1, \;\; m \in T_2$
- We compute m as sim(n,m) is maximum
- $\text{stead}(n) = \text{sim}(n, m) + \mu \times \text{stead}(m)$. With T_0 as base case and $\mu \in \Re$.

So we compute the global steadiness of a Tree T by:

$$\text{Persistence}(T) = \frac{\sum_{n \in T} \text{stead}(n)}{|T|}$$

Outline

Data Set

Aggregation Window: 1 Week Time Length

- Macro: Up to 52 weeks from 2011-04-23 to 2012-06-30 (662 K)
- Micro: 10 weeks maximum

Data Set

Aggregation Window: 1 Week Time Length

- Macro: Up to 52 weeks from 2011-04-23 to 2012-06-30 (662 K)
- Micro: 10 weeks maximum

Malicious data

- Time: Periodically, Steady

Data Set

Aggregation Window: 1 Week Time Length

- Macro: Up to 52 weeks from 2011-04-23 to 2012-06-30 (662 K)
- Micro: 10 weeks maximum

Malicious data

- Time: Periodically, Steady
- Proportion: 0.1%, 1% and 10%

Data Set

Aggregation Window: 1 Week Time Length

- Macro: Up to 52 weeks from 2011-04-23 to 2012-06-30 (662 K)
- Micro: 10 weeks maximum

Malicious data

- Time: Periodically, Steady
- Proportion: 0.1%, 1% and 10%
- Source: Blacklists (Exposure, WOT) 175K

Aggregation α: 2%

Steadiness Distribution

Distribution of Local Steadiness (Leafs)

- More than 50% of malicious nodes have less than 0.7 of steadiness

Steadiness Distribution

Distribution of Local Steadiness (Leafs)

- More than 50% of malicious nodes have less than 0.7 of steadiness
- Less than 20% of malicious nodes have more than 0.85 of steadiness

Steadiness Distribution

Distribution of Local Steadiness (Leafs)

- More than 50% of malicious nodes have less than 0.7 of steadiness
- Less than 20% of malicious nodes have more than 0.85 of steadiness
- Less than 40% of normal data have a steadiness of 0.8 or less.

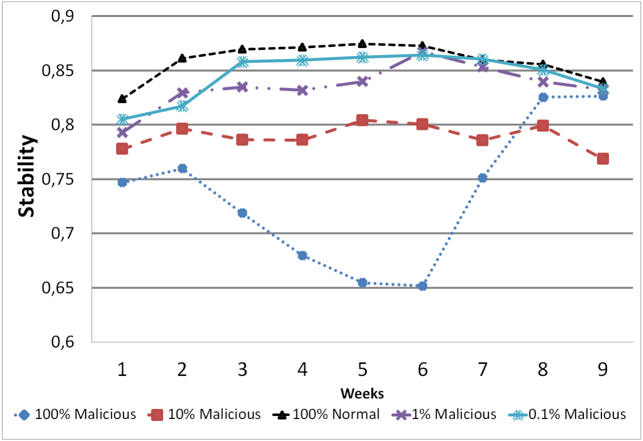
Steadiness Distribution

Distribution of Local Steadiness (Leafs)

- More than 50% of malicious nodes have less than 0.7 of steadiness
- Less than 20% of malicious nodes have more than 0.85 of steadiness
- Less than 40% of normal data have a steadiness of 0.8 or less.
- Only 10% of normal data have a steadiness of less than 0.5

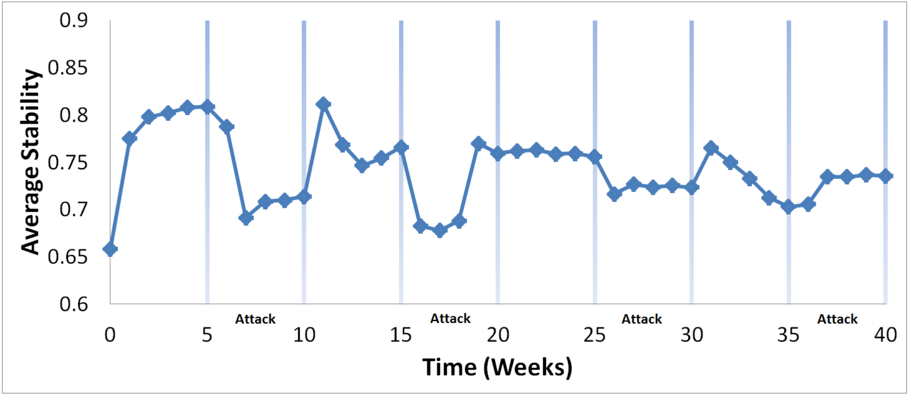
Microscopic observation

Malicious domains causes a drop on average steadiness



Macroscopic observation

Malicious domains causes a drop on average steadiness: Macro



Contributions

Summary

- A methodology for assessing DNS - IP association time frame was proposed.

Contributions

Summary

- A methodology for assessing DNS - IP association time frame was proposed.
- Reduced the scale of data, helpful in the context of network security. (From 80K nodes to 2K with $\alpha = 2\%$)

Contributions

Summary

- A methodology for assessing DNS - IP association time frame was proposed.
- Reduced the scale of data, helpful in the context of network security. (From 80K nodes to 2K with $\alpha = 2\%$)
- Definition of steadiness metrics for a local and global scope was introduced.

Summary

- A methodology for assessing DNS - IP association time frame was proposed.
- Reduced the scale of data, helpful in the context of network security. (From 80K nodes to 2K with $\alpha = 2\%$)
- Definition of steadiness metrics for a local and global scope was introduced.
- Evaluation using real data and during several time frames. Validation of the metrics

SnT

Outline

Motivation

Why application-awareness is important at network level?

- Better Network Management in terms of applications
  - Routing
  - Policy implementation: QoS, Security, etc
  - Resource Allocation
  - Developers
- Network Monitoring
  - Per Application
  - Per Instance / User
  - SDN Potential

Outline

State of the Art

SDN Background

Software-defined networking (SDN) decouples the data and control planes

- Flexibility
- Traffic Efficiency
- Open Flow as de-facto standard

Still a lack of global application perspective

State of the Art

SDN Background

Software-defined networking (SDN) decouples the data and control planes

- Flexibility
- Traffic Efficiency
- Open Flow as de-facto standard

Still a lack of global application perspective

Application Awareness

Network Responsive

The applications convey to the network bandwidth requirements

Application Responsive

The applications include the network status for making decisions

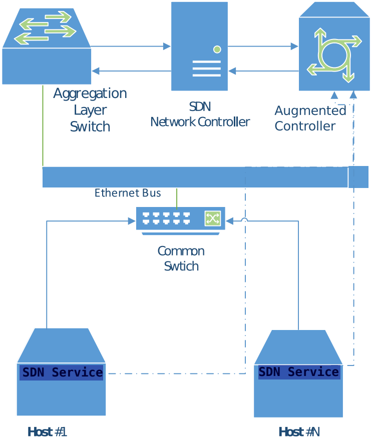
Conventional Networking

Given the traffic demand matrix is possible to optimize the resources

Outline

Proposed Architecture

Architecture

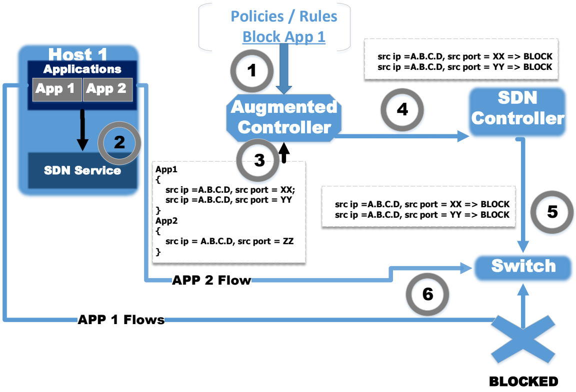


Components & Connector

- Augmented Controller (AC)
  - Connects to SDN Controller and SDN Services
  - Has a global view of flow/application
- SDN Service:
  - Runs at host level attached to applications
  - Flow information per application to be sent to the AC

SnT

Example of Application-Aware Policy



Outline

Application Level Flow Management: Evaluation

Impact on performance

- We evaluated using several applications such as: Iperf, Apache2, Links, Pidgin
- We registered CPU & Memory at Host Level
- Networking Metrics such as:
  - Throughput
  - % Packet Loss
  - Bytes Transferred

We studied the impact on computing resources consumption and collateral network traffic

Application Level Flow Management: Results

Impact on CPU consumption

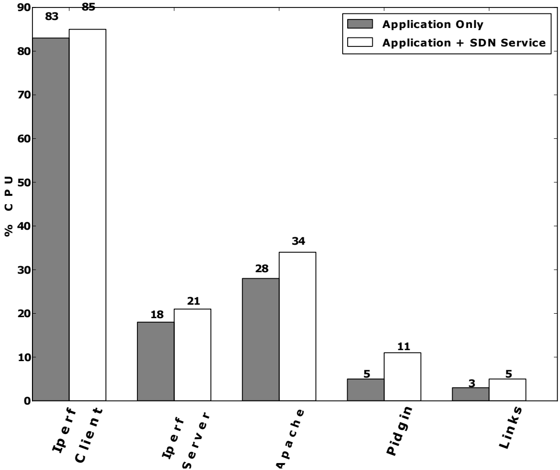
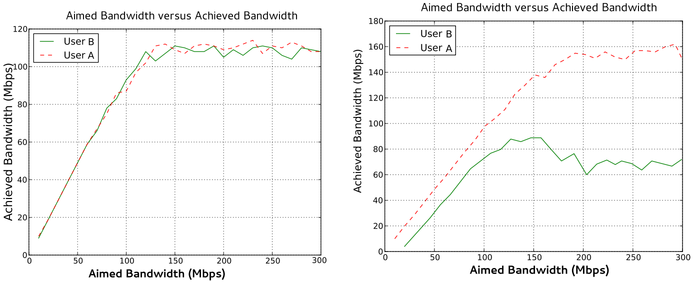


Figure : Example of the impact on performance upon SDN Service usage

Application Level Flow Management: Results III

Network Performance



(a) SDN Service Inactive (b) SDN Service Activated

Figure : Bandwidth Aimed vs Bandwidth Achieved

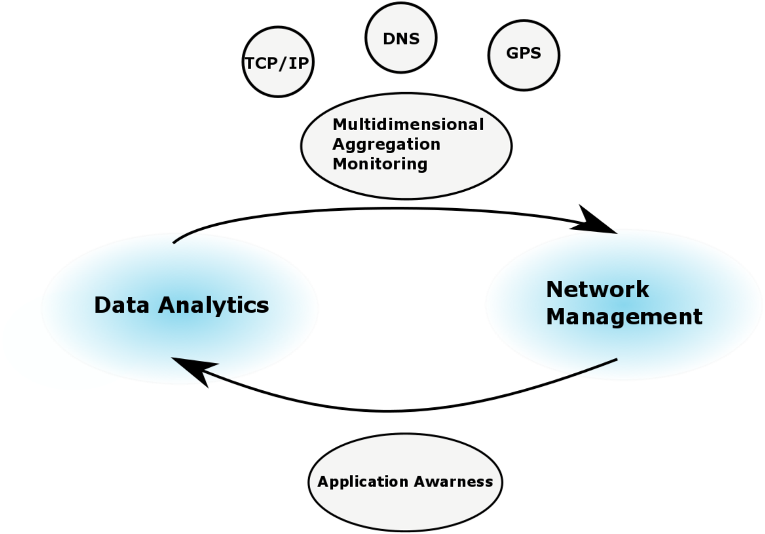
Conclusions

Application Awareness

- Validated an Application-awareness approach is feasible
- Small impact regarding computing & network overhead
- Evaluation with multiple types of applications
- Implemented a Proof of concept
- Extensible API

Outline

Summary


DNS
TCP/IP
GPS
Multidimensional Aggregation Monitoring
Data Analytics
Network Management
Application Awarness

Summary Cont'd

Aggregation

- Reduced the scale of data
- Preserved natural hierarchy of data (semantics)
- Minimized the information loss
- Efficient Algorithms

Application Awareness

- Network Awareness
- Software Defined Networks
- Proof of Concept & Test Bed
- Examples: Map Reduce, QoS, Security

Future Research Perspectives

Security: Network Monitoring

- Extension of other domains
- Aggregation based on events
- Aggregation per dimension

Network Management

- Multi tenant Data Centres
- Monitoring per application / instance
- Scheduling of distributed applications
- Logging

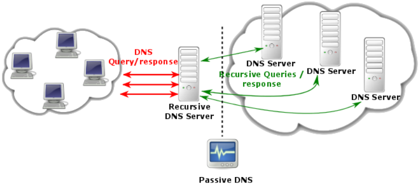
Research Experience

- MaM *Efficient multidimensional aggregation for large scale monitoring, LISA 12, USENIX*
- DNS *DNS Stability Evaluation using Multi-Dimensional Aggregation, LCN 2013*
- Vehicular Networks *Tracking Spoofed Locations in Crowd-Sourced Vehicular Applications, NOMS 2013, IFIP/IEEE*
- Assessing artificially caused congestion on urban scenarios: A case study on Luxembourg sumo traffic, in 3rd GI/ITG KuVS Fachgesprach Inter-Vehicle Communication, 2015.
- Mobile Security *String-Based Android Malware Detection SecureComm '14, EAI*
- Network Security through Software Defined Networking: a Survey *IPTCOMM 14*
- Application-Level Flow Management For SDN-Based Cloud Infrastructure IM 2015, Work in Progress
- Book Chapter for Networking on Big Data Application In Collaboration with University of Waterloo

Outline

Passive DNS

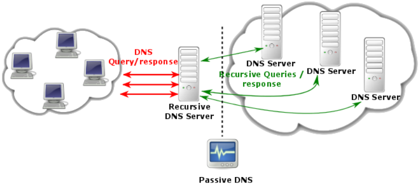
A Passive DNS DB contains:



- Where did this domain name point to in the past?

Passive DNS

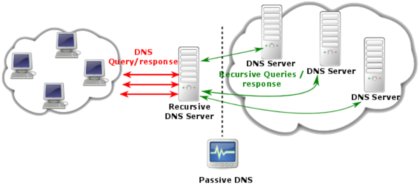
A Passive DNS DB contains:



- Where did this domain name point to in the past?
- What domain names are hosted by a given nameserver?

Passive DNS

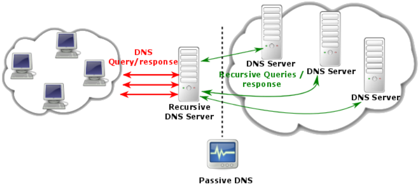
A Passive DNS DB contains:



- Where did this domain name point to in the past?
- What domain names are hosted by a given nameserver?
- What domain names point into a given IP network?

Passive DNS

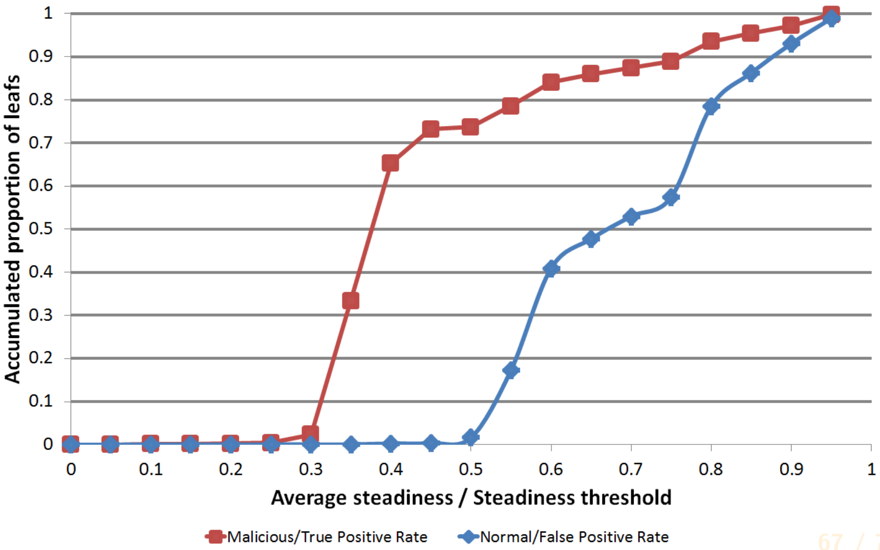
A Passive DNS DB contains:



- Where did this domain name point to in the past?
- What domain names are hosted by a given nameserver?
- What domain names point into a given IP network?
- What subdomains exist below a certain domain name?

Threshold based detection

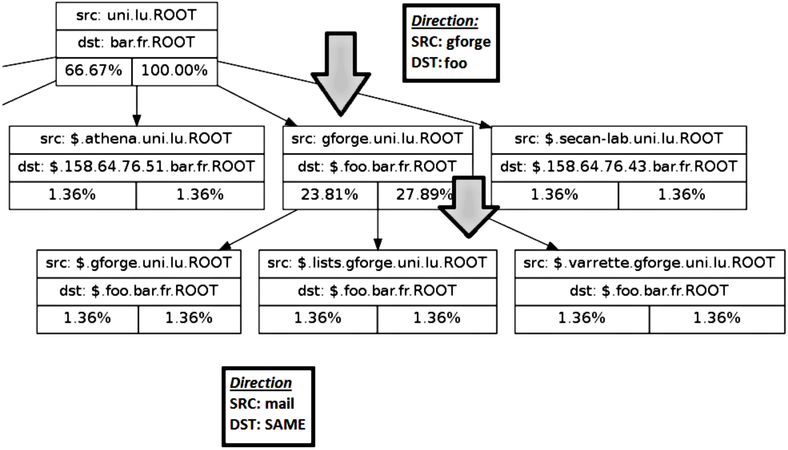
Accuracy: Stability as metric for filtering malicious domains



Data Structure (II)

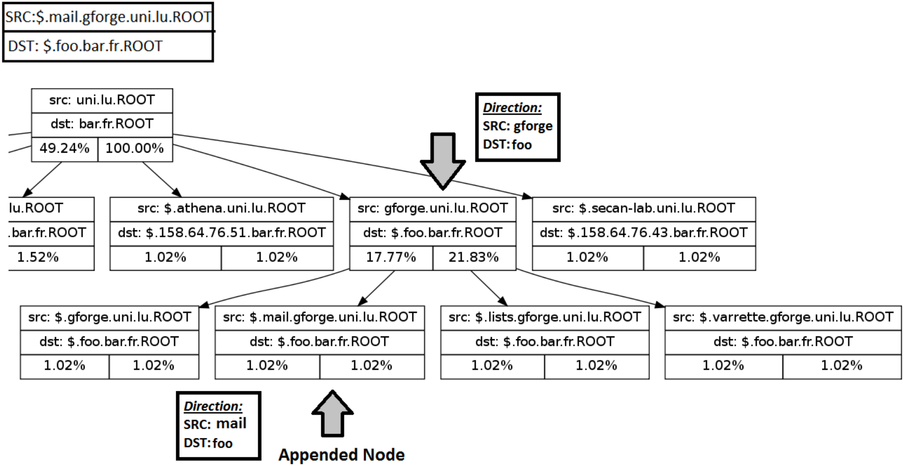
Node Insertion (Partial Match)

SRC:$.mail.gforge.uni.lu.ROOT
DST: $.foo.bar.fr.ROOT



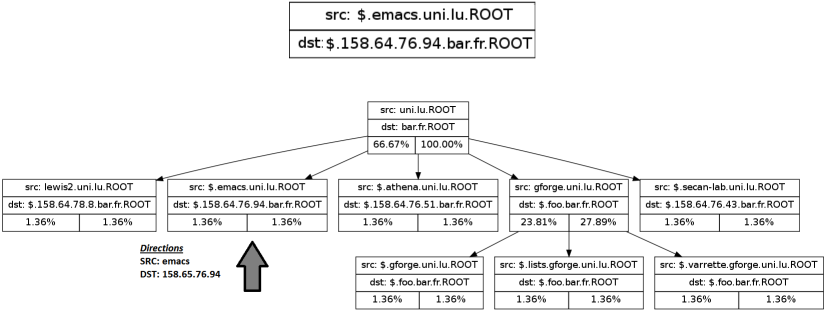
Data Structure (II)

Node Insertion (Partial Match)



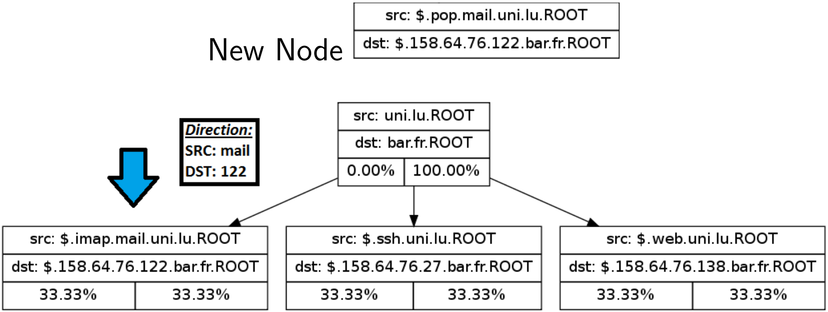
Data Structure (II)

Node Insertion (Perfect Match)



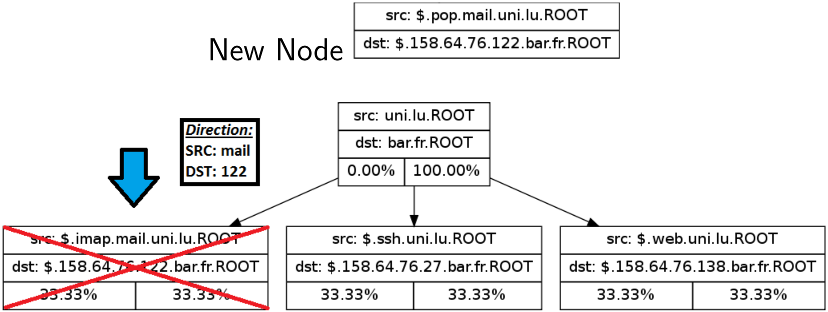
Data Structure (II)

Node Insertion (Branching Point)



Data Structure (II)

Node Insertion (Branching Point)



Data Structure (II)

Node Insertion (Branching Point)