TroubleMiner: Mining Trouble Tickets

Amélie Medem K.,
Marc-ismael Akodjenou
Renata Teixeira

CNRS and UPMC Paris Universitas, Univ Paris 06
LIP6 Laboratory

What is a trouble ticket?

Identity & Time

Problem Description

Actions

Ticket 20070904_2

Ticket Number:	20070904_2	**Ticket State:**	CLOSED
Ticket Opened:	2007-09-04 20:44	**Ticket Closed:**	2007-09-10 10:01
Ticket Description:	Router swiEL2.switch.ch hiccup due to DDoS		

Problem Description:

Last Saturday (01 September 2007) around 01:51, our router swiEL2.switch.ch suffered a period of CPU overload which caused some routing processes to fail. The overload was due to an intense flood of traffic sent to one of the router's interfaces, probably as part of a denial-of-service attack against a host at EPFL.

From 2007-09-01 01:51 until 2007-09-01 01:54
Impact: no more redundancy
Sites/Services: EPFL, IMD

Impact

Actions:

2007-09-10 10:01
Since the issue hasn't surfaced anymore, we close the ticket. We will continue to study methods to protect our routers' "controle-plane" processing against such traffic.

2007-09-04 16:00
An analysis of the traffic on our border routers during the time of the incident showed a flood of tiny packets against an inactive TCP port of our router, which must have caused extreme overload of the router's CPU.
We are considering ways of protecting the router against this kind of traffic.

2007-09-03 12:49
EPFL noticed that there had been an outage of their BGP session to swiEL2.switch.ch. It was suspected that this was a reoccurrence of the problems in May, see ticket 20070510_1. Our first analysis showed CPU overload, but the reason wasn't clear.

For all questions about this ticket, please send mail to noc@switch.ch
or call +41 44 268 15 30.

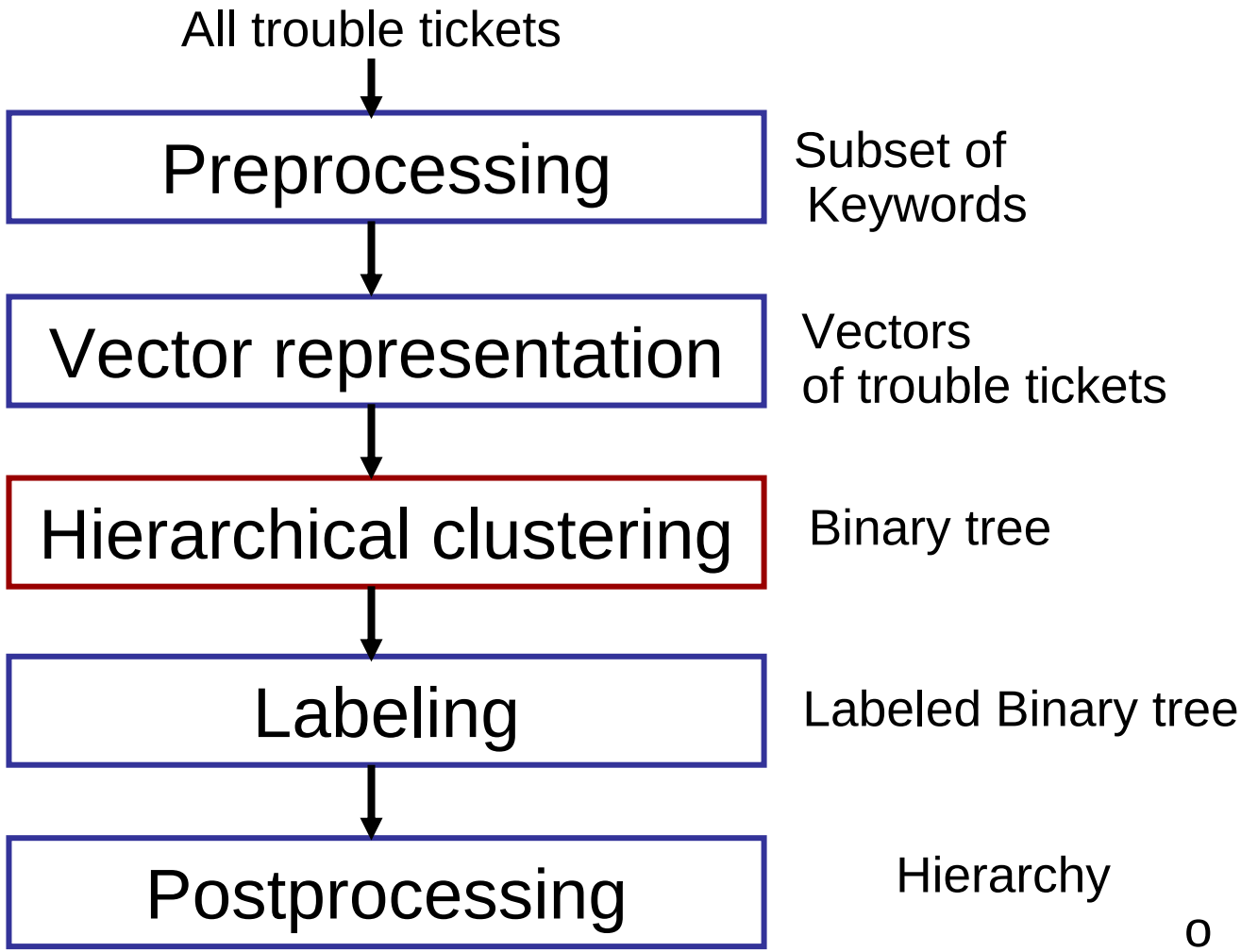
Trouble tickets contain strategic information

- Statistical trend analysis
  - Most frequent events
- Validation of other tools
  - Causes of routing changes or traffic anomalies
- Help troubleshooting network incidents
  - Use past experience to guide the resolution of similar incidents in the future
- But, trouble ticket analysis is hard
  - No standard in the description of trouble ticket
  - Description is written by hand
  - The number of tickets can be very high

Our goal

- Automatic processing of trouble tickets

TroubleMiner
Mining the content of trouble tickets



Keyword selection

- Process all words in the Subject and Description
- Stem words: eliminate plurals, suffixes
- Remove of low frequency words: >50% of words in one trouble ticket
- Remove of irrelevant words (like stop-words)
- Correct of mistakes and group similar words in equivalence classes

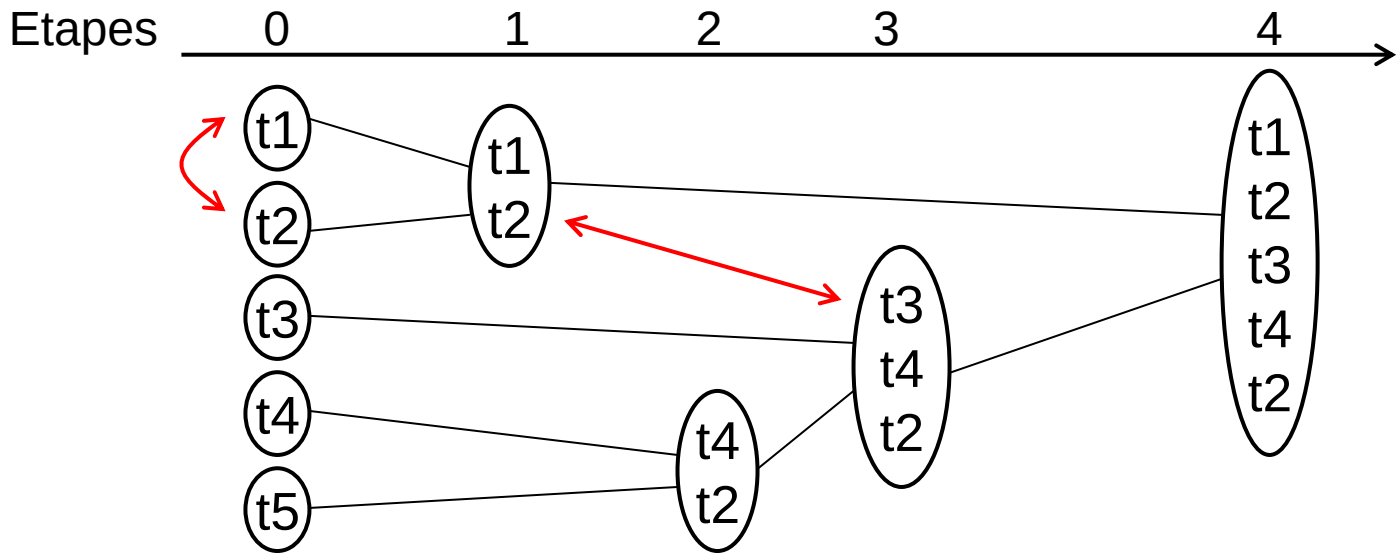
tt1: Scheduled mainten ance fiber splic e works between Domodossola …
tt2: Planned mainten ance softwar e upgrad e on the primari access router…
tt3: Heavi traffic caused an overload of multicast and crash our router …

A vector per trouble ticket

tt1 label: (mainten, fiber, splic)
tt2 label : (mainten, softwar, upgrad, access, router)

- Problem: all keywords have the same importance (weight) in trouble tickets
- Weight=Document Frequency (DF)

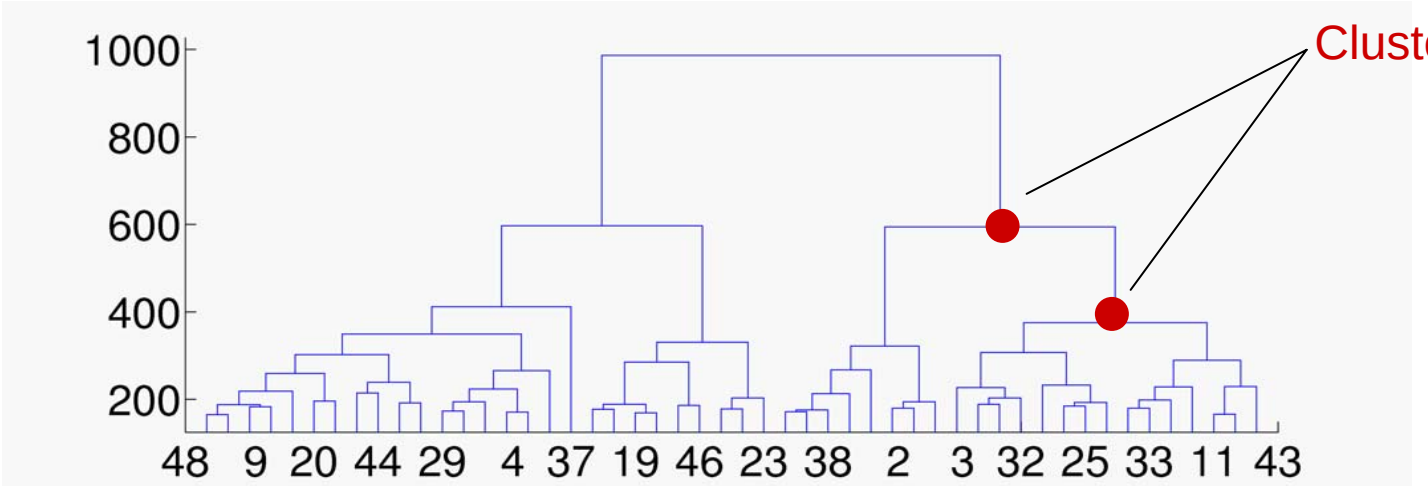
Hierarchical clustering



- Compute all pairs of distance between trouble tickets
- At each step, merge the 2 closest trouble tickets or clusters

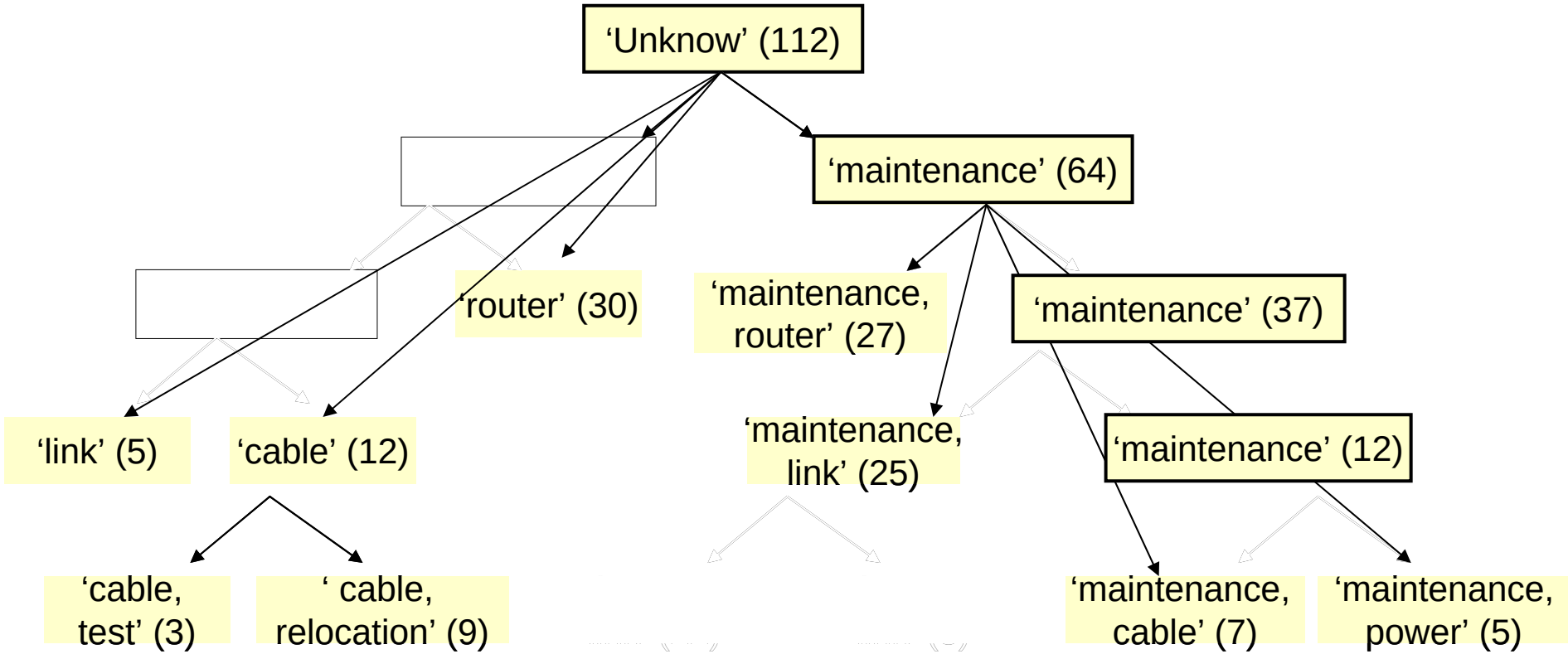
- Need 2 metrics : Distance between trouble tickets (Euclidean, Cosine) and Similarities between clusters (average, single, complete link)

Labeling the binary tree



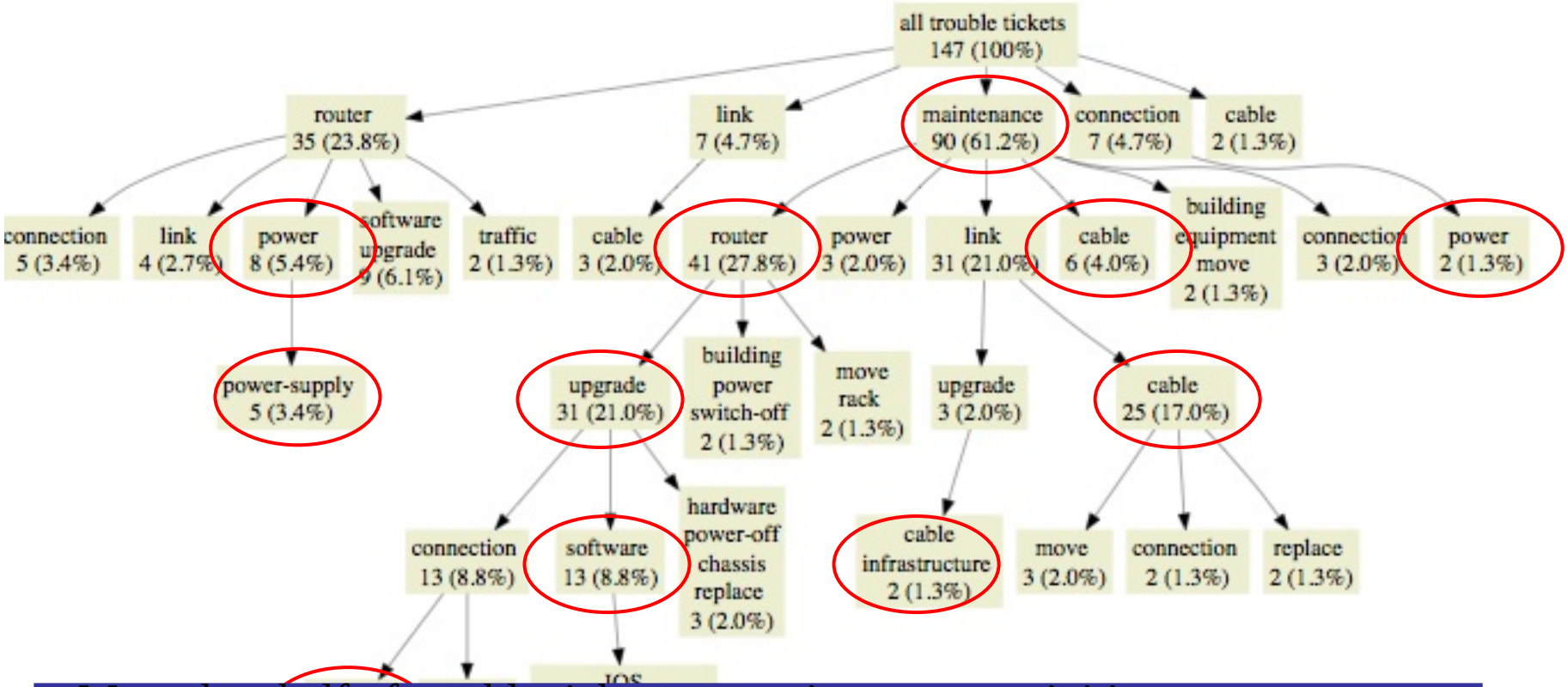
- Label from leaves to the root
- Label of a trouble ticket is its set of keywords
- Label of cluster = intersection of keywords of all children in subtrees

Pruning the binary tree



- Merge nodes with the same label
- Create new nodes when necessary

Characterization of trouble tickets Switchlan case studies



- More than half of trouble tickets are maintenance activities
- Maintenance activities are mostly router software upgrades and cable work
- Power outages and routing errors are less frequent

Conclusion

- TroubleMiner :
  - Meaningful hierarchy from the disorder of trouble tickets
  - Model of trouble tickets: vector of keywords
  - Algorithm to transform the binary tree into a labeled hierarchy.
- Perspectives:
  - Use the TroubleMiner as a basis to assist in the troubleshooting procedure.

- Thank you!