Evaluating Internet Background Radiation Detector Rules

Diploma Thesis

Department of Computer Science
University of Applied Science Rapperswil

Spring Term 2011

Author(s):	Mathias Vaterlaus
Advisor:	Prof. E. Glatz
Project Partner:	ETH Zürich
External Co-Examiner:	Roberto Pajetta

Statement against plagiarism

Statement

I hereby declare,

- that I have done the work on hand without any help of a third party, except the help, which is explicitly mentioned in the conceptual formulation or the one which was aggred in wirtten form with the tutor.
- that all of the resources used are mentioned and and documented according to the scientific citation rules.

Rapperswil, the 06.05.2011

Mathias Vaterlaus

0. Document Information

0.1. Change History

Date	*Version*	*Changes*	*Author*
07.03.11	0.1	Initial Version	mvaterla
09.03.11	0.2	Start of research Documentation	mvaterla
14.03.11	0.3	Prestudy described	mvaterla
21.03.11	0.4	Prestudy Frequent Item Set Mining	mvaterla
24.03.11	0.5	„Applying FIM“ documentation started	mvaterla
12.04.11	0.6	FIM corrections	mvaterla
28.04.11	0.7	Sign statistics commented	mvaterla
30.04.11	0.8	FIM analysis over whole period commented	mvaterla
04.05.11	0.9	FIM results over reference intervals commented	mvaterla
06.05.11	1.0	Release	mvaterla

0.2. Index

1. Introduction

1.1. Initial Situation

A significant amount of the traffic in the Internet is non productive traffic. This is unwanted traffic because it is caused by scanning (D)DoS attacks, misconfiguration and other. Observing this traffic and analyzing its long term evolution is important to observe statistical trends in malicious activities and to combat their occurrence. Because of its ubiquitous nature and the variety forms of appearance, such traffic is referenced to as Internet Background Radiation.

A new detection technique was developed at the ETH Zürich and implemented in an Internet Background Radiation detector. This detector is written in C++ and is used to analyze flow data. The first step of this technique is to separate the whole traffic into uni- and bidirectional flows. In a second step, a defined rule set is applied to the unidirectional flows to identify the different types of it. Long term analysis have been made which brought up some interesting results. What is missing at this time is a in-depth evaluation of the causes being collected with these detection rules to optimize the detector.

1.2. Goals

- Evaluate, if the IBR Detector classifies the inspected one-way flows correctly.
- Out of the periods classified by the IBR Detector, some show a significant peak in assigned flows. The second goal is to find the causes of the peaks.
- Investigating the effectiveness of rules for each class having more than one rule which matches flows to it.

1.3. Approach

For reaching the defined goals, various analysis are made. First, all flows and their according signs of the inspected intervals of a class are extracted from the cflow files and their sign-set files. Both resulting output files are then inspected with the frequent item-set mining (FIM) technique. The second analysis counts all occurrences of each sign over an inspected interval. The third analysis compares all flows of a class against the flow item-sets found in the same interval of the same class. If the flow matches the flow item-set, all sign occurrences belonging to this flow are recorded.

1.4. Document overview

Chapter 2 introduces the basics to perform this analysis. This chapter covers all protocol information important to identify the causes of observed flow item-sets. After that, the related work is covered in chapter 3. Chapter 4 introduces the FIM theory as well as the CISCO netflow format. The FIM analysis over randomly chosen continuous three hour intervals are described in Chapter 5. Chapter 6 and 7 cover the FIM analysis of reference intervals of a non peak period and the FIM analysis over the whole peak periods. Sign statistics are described in Chapter 8. The sign statistics cover the occurrences over the whole inspected interval as well as the statistics over all occurring signs per flow item-set. The next chapter, chapter 9, describes the final results of this diploma thesis completed by the following bibliography in chapter 10.

2. Protocol Basics

2.1. Introduction

In this chapter, the basics of the most important protocols are described. It documents the most common transmission protocols such as TCP and UDP, the underlying IP protocol and the ICMP protocol. In addition, some other protocols are introduced, which might be interesting for identifying some of the one way flows.

2.2. IP

Nowadays, the Internet uses IPv4 as standard protocol for routing data through it. An IP packet does encapsulate layer 4 transport protocols and is used by routers to direct the data through the network. The IP Header includes Information about the sending and receiving IP address and will be discussed in the next section. Note that in this document, only IPv4 will be mentioned. IPv6 will not be covered.

2.2.1. IP Header

The IP header includes several fields. This section will only describe the most interesting ones for the projects task. For more information about IP and its details, take a look at the literature list in the appendix of this document. Illustration 1 shows the complete IP header. Header fields which are of interests for the FIM analysis are described in table 1.

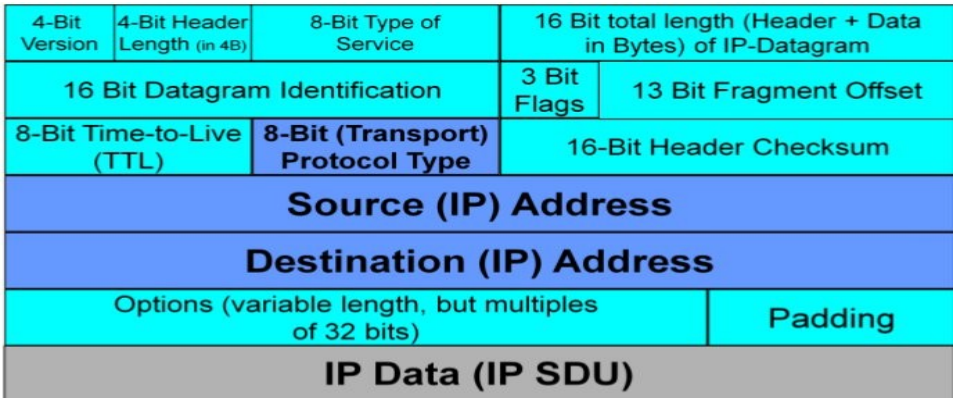

Illustration 1: IP Header [1]

Field	Description
Version	Defines the version of IP. (IPv4, Ipv6)
Type of Service	ToS or Type of Service describes several QoS options for the traffic at router level. This includes minimizing delay or maximizing throughput among other. The following settings of the flags define the different options: • 0000: normal service • 0001: minimize monetary cost • 0010: maximize reliability • 0100: maximize throughput • 1000: minimize delay
Flags	The IP header defines 3 flags. The first one is reserved and should be 0. Second is the Don't Fragment (DF) flag, which indicates if it is permitted to split the datagram. The last flag is the mark segment (MF) flag, which is set if this datagram is a fragment and there are more to come. The last datagram of the larger packet has not set this bit, to indicate, its the last one.
Time to live	TTL sets, how long the packet is valid. The counting is done on a per hop basis.
Transport protocol type	Defines the protocol encapsulated in the IP datagram. The most common protocol types are the following: • ICMP (number 1) • TCP (number 6) • UDP (number 17) Detailed lists of protocols are available on IANA [2] and wikipedia [3].

Table 1: IP fields useful for FIM analysis

2.2.2. Conclusions

Taking a look at the IP datagram header gives some possibilities to identify IBR traffic. The first is to observe, if malformed packets exist. This can be done by the examination of the flags, if they are set appropriate. Also the header checksum field could give a hint, if a package is malformed or manipulated. This can help in identifying malicious traffic generated by virus, worms, Trojans or some uncommon network scanners, which are poorly designed.

The TTL header field can be used for identifying trace route commands which are implemented by sending ping messages with increasing the TTL header field from 0 until the ping is being answered.

We also need to take a close look to the type of service field, because it might give some applications which use this filed for a higher quality of service. In DoS attacks, this could be used to give the flooding packets a higher priority, so they will be handled before the normal traffic.

The protocol type field is used to identify the transport layer protocol. The generation of ICMP messages could be caused by a protocol number not supported, or by a non existing protocol number.

2.3. TCP

TCP is a connection oriented protocol, which indicates that a connection setup has to be made, before data can be exchanged. The connection setup is called a 3 way handshake and is controlled by the flags defined in the header of TCP. This section covers the fundamentals of TCP, including connection setup and tear down, important fields of the TCP Header as well as the explanations of the flags and the different state a TCP connection can be in.

2.3.1. TCP Header

In this section, interesting fields will be described in table 2. The header flags are described in the section 2.3.2 TCP Header Flags, because they are very important to determine the state of a connection. A sample visualization of the TCP header can be viewed in illustration 2.

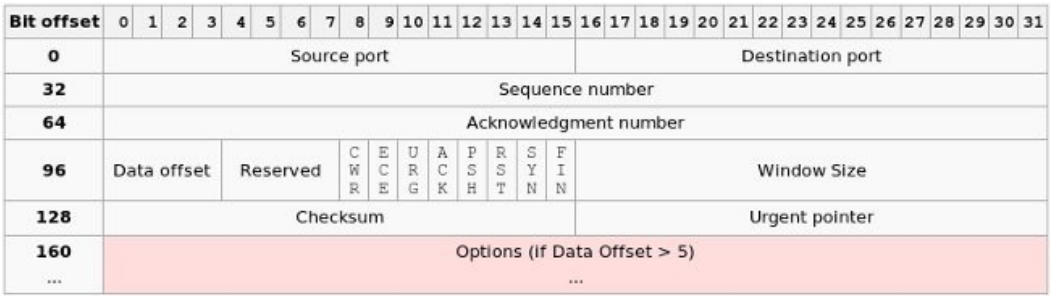

Illustration 2: TCP header [4]

Field	Description
Source port	The source port determines, from which port the packet was sent.
Destination port	This field describes the destination port, to which a packet is sent. This can help to identify the service which is running on the inspected host.
Sequence number (SN)	Within the connection establishment (SYN flag set), SN tells the other party the starting SN. The other party replies with a ACK (SN + 1), which is the starting SN of the following data segments. During the communication, sequence numbers are used to handle the incoming segments in the right order. The SN refers to first byte of data in segment.
Ack. number	This field contains the sequence number of a previous package. It tells the sender of the package, that the segment with this sequence number has been transmitted successfully. If it is during the 3 way handshake, the acknowledge number is (SN +). It is possible to acknowledge multiple segments (cumulative ACK or Selective ACK). Usign cumulative ACK's, only the last received segment is acknowledged. With SACK, a maximum of four blocks can be acknowledged, because of its maximum field width. To reduce Sack's, it should only be used to acknowledge the most recent data received.
Options	The options field includes some special attributes to specify the behavior of TCP transmission. Interesting attributes are the selective acknowledgments, which can identify packet loss. The window scaling attribute can help to identify malformed packets when this attribute is set an no SYN flag is set. The window scaling attribute can not be changed and is set during handshake.

Table 2: TCP header fields

2.3.2. TCP Header Flags

This list doesn't include the new flags (CWR and ECE) defined in RFC 3168 [5], because they are optional. The flags are only set, when both stations involved in the communication are capable of ECN. Only their existence is important, because they borrow two bits of the header field „Reserved“, which is important for analyzing the flags of TCP. Table 3 shows a listing of important flags to determine the state of a TCP connection.

Flag	Description
URG	Marks this packet as urgent. This is used by some applications to react on Ctrl + C commands.
ACK	Acknowledge: Is used to acknowledge a received packet.
PSH	Push: Asks to push the sent data immediately to the Application it belongs to.
RST	Reset: Resets the connection.
SYN	Synchronization. This flag indicates the attempt of a connection.
FIN	Signals to the communication partner to tear down the connection.

Table 3: describing TCP header flags

2.3.3. Connection establishment and tear down

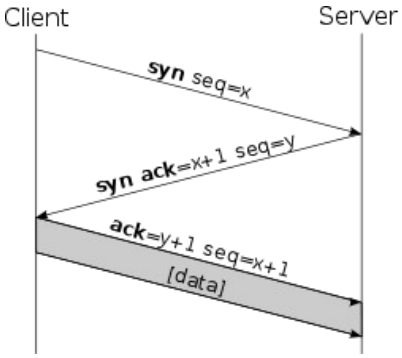

Illustration 3: TCP connection establishment [6]

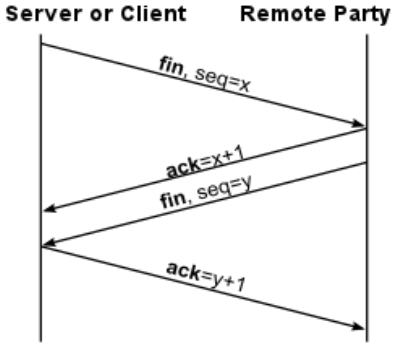

Illustration 4: TCP connection tear down [6]

Illustrations 3 and 4 show the typical TCP connection establishment and tear down.

2.3.4. TCP Connection States

The connection states of TCP can be divided in to two sections: The connection setup and the connection tear down. TCP defines 11 states in total, which will be described in this section, structured into the two sections. The information to describe the states are contained within the book “Inside Network Perimeter Security” [7] in chapter 3.

Connection establishment

- CLOSED:
  A “non-state” that exists before a connection actually begins.
- LISTEN:
  The state a connection is in, when a server listens for requests on a given port to start a connection. This is the true starting state of a TCP connection.
- SYN SENT:
  The state when a machine has sent a SYN package to the target host and is awaiting the proper SYN ACK reply from the target host.
- SYN RECEIVED:
  Defines the state a host is in after receiving a SYN packet and sending the proper SYN ACK reply to the initiating host.
- ESTABLISHED
  In this state, the connection is established. The initiating host goes in to this state after receiving a SYN ACK packet and the replying host changes to this state after the lone ACK packet from the initiator.

Connection tear down states

- FIN WAIT 1:
  The state a connection is in after sending a initial FIN packet asking to close the TCP connection.
- FIN WAIT 2:
  A host is in this connection state after receiving the ACK of its initial FIN and thus waiting for the final FIN packet from its opponent.
- CLOSE WAIT:
  This state is entered after a host receives the initial FIN packet and responds with an ACK packet.
- LAST ACK:
  State of the host that sent the last FIN to normally close the connection while waiting for an ACK packet.
- TIME-WAIT:
  Final connection state of the host who initiated the connection. After sending the final ACK packet to acknowledge the last FIN packet, it has to wait a given time period before changing to state “Closing”, so the other host has enough time to receive the final ACK packet.
- CLOSING:
  When using nonstandard simultaneous closing, a connection goes into this state after receiving an initial FIN and sending an ACK. If it receives an ACK for its FIN, the connection state changes to TIME-WAIT

2.3.5. Conclusions

Analyzing the header fields of TCP, especially its flags, can be used to identify the states a TCP connection is in. Within the state analysis of the connections, it is possible to filter flows which might be an attack due to transmission errors or incorrect state behavior.

Malformed packets can be identified when the window scaling option changes during the connection. The window scaling option in the TCP header can only be set during connection setup when sending the SYN packet.

Analyzing the selective and cumulative acknowledgment options gives information about retransmitting of frames. Packets headers could have been corrupted during the transaction which leads to retransmitting the packets.

During handshake it is sent after receiving the initial SYN packet. If a communication exists between two endpoints, RST signalizes to reset the connection immediately. Therefor the RST flag indicates connection attempts to a non existing service. This can be very helpful in determine misconfiguration.

2.4. UDP

UDP is a connectionless transport protocol and does not provide a connection setup mechanism. The sender of datagram does only need to know, on which port the receiver of the datagram is expecting it. Applications, which are built upon UDP must implement transmission error- or flow control handling by them selves. The design of UDP and its header generates lesser overhead than TCP and allows faster data exchange.

2.4.1. Conclusions

As there is no further information contained in the header of UDP datagrams than source- and destination port numbers, as well as the length of the datagram and a checksum, information about possible suspicious traffic must be gathered somewhere else.

The first possibility is to inspect the data of the UDP datagrams and identify known services or protocols. Secondly the IP header field ToS could lead to a deeper understanding of the analyzed traffic because some applications, for example multimedia applications, use this field to maximize throughput or to reduce latencies.

2.5. ICMP

ICMP is based on the IP protocol and is used to notify other devices of failures in particular machines. It is also an integral part of IP which meas, that every IP module must implement ICMP. The protocol does not make IP reliable, it is only used to report errors of sending data to a particular host. Replies are never sent as response to a datagram addressed to a uni- or multi cast address. They will only be generated, if a host is uniquely identifiable. In addition, ICMP can be used to react on datagram processing errors but this is an optional feature and its availability is depending on the corresponding implementation.

2.5.1. Implementation in IP

As mentioned before ICMP is carried in the payload of an IP datagram. To identify the IP datagram as an ICMP message, the protocol type field is set to the value 1 and the type of service bit is always set to 0. ICMP messages have three header fields (ICMP type, ICMP code, checksum) and payload. To identify which message has caused the ICMP datagram, the generator of the ICMP datagram adds the first 8 bytes of the message which caused the generation.

2.5.2. ICMP types and codes

ICMP provides a lot of different message types an codes. Illustration 5 shows only the most common types and according codes of ICMP messages. For a detailed list of all ICMP types and codes, please refer to IANA assignment [8].

type	code	description	application
0	0	echo reply (ping reply)	used in probes
3	0	destination network unreachable	improves performance
3	1	destination host unreachable	improves performance
3	2	destination protocol unreachable	improves performance
3	3	destination port unreachable	improves performance
3	4	fragmentation needed but "don't fragment bit"	
5	0	redirect for network	improves performance
5	1	redirect for host	
8	0	echo request (ping request)	used in porbes, traceroute
11	0	time exceeded: TTL equals 0 during transit	used in traceroute
12		Parameter Problem	reports header errors
13		Timestamp	used in probes
14		Timestamp Reply	
17	0	subnet mask request	local use
18	0	subnet mask reply	

Illustration 5: Common ICMP types and according codes [1]

2.5.3. Echo: request and reply

The two echo messages are special to ICMP because they are not sent based on an appeared error. First the request is generated on behalf of the known command ping. If the request reaches the destination IP address, the receiving host changes the type of message to echo reply and sends the packet back to originating IP address. It is possible, that a policy on the host exists, that forbids to reply on ICMP requests, or that a firewall blocks such requests.

2.5.4. Conclusions

To assign a outgoing ICMP message to the corresponding packet, the first 8 bytes of the ICMP payload can be analyzed. These first 8 bytes contain the header of the packet, which has caused the generation of the ICMP message.

The occurrence of a big amount of some ICMP messages (for example type 3 code 1) can help identifying an ongoing attack on a machine. If the machine itself generates a lot of ICMP messages to real or spoofed IP addresses the same machine is being scanned or attacked. If ICMP messages of the before mentioned type and code are incoming, it most probably is an ongoing attack on the host, who generates the messages.

Echo request and replies of ICMP occurring at different hosts on the same network or subnet in a short range of time is most probably a network scanner trying to identify the on-line hosts in a network. The difficulty in this analysis method is to correlate the ICMP replies of all hosts. If an echo responses from a particular host is followed by some ICMP messages saying the destination port is unreachable or destination host is unreachable. This could be a clue to an upcoming attack against this host.

For echo requests an option exists, which records the route being taken to the target host. Because of the maximum length of the header options field, only nine hops can be recorded. Many known systems exists, which do not support this option or are just ignoring it.

Many programs use MTU path discovery to figure out, how large packets can be, so they don't need to be fragmented. MTU path discovery tools set the DF (don't fragment) bit in the IP header an start probing with different payload lengths to figure out, which is the largest payload the path to a target supports.

Packet header errors generate parameter problem ICMP messages. If the inspected host generates a big amount of these messages, this can be a sign for misbehaving application or of a threat caused by malicious software, for example viruses and worms.

To identify the most common ICMP types and their codes, it would be useful to do a frequent item set mining over all ICMP traffic.

3. Related Work

3.1. Visualizing Host Traffic through Graphs [9]

3.1.1. Short overview

This paper introduces a new technique of visualizing flow data in a simple an well organized way. The tool HAPviewer implements this technique and visualizes flow data as 5-partite graph. The presentation of the traffic is based on the key attributes of the Berkeley socket model. Lesser important links are pruned so each of the socket attributes (local IP, protocol, local port, remote port, remote IP), which are represented as nodes, can be placed on the according partition. In this document, only the most important parts are discussed. If you are interested in the whole paper, a link is attached to footnote 6. How such a graph looks like is shown in the image of section 3.1.3 “Host role summarization”.

3.1.2. Flow classification and filtering

HAPviewer assumes, that any productive traffic sent from a host is answered by its communication partner. So it distinguishes between uni- and bidirectional traffic. Unidirectional traffic is in most cases suspicious and needs further investigation. Bidirectional traffic is marked with a thick line and arrows at both ends, unidirectional traffic is marked with a single arrowed line pointing in the direction of the data flow.

All unidirectional flows are divided into two categories. If the same hosts have communications established, the unidirectional flows, which are not answered are marked as failing connections and are shown with a green line. All other unidirectional traffic is marked in red.

3.1.3. Host role summarization

The host role summarization done by HAPviewer includes the roles shown in illustration 6 [10]. Identified host roles are summarized to a single flow. The summarization done by HAPviewer is shown in form of squares used to display the nodes in stead of the normal ellipses. The summarization process is performed in the alphabetical order of the role definitions.

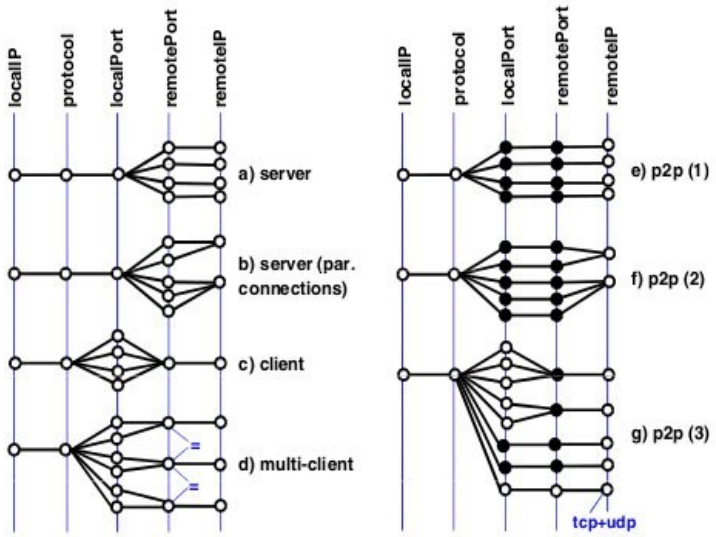

Illustration 6: Image 1: Host roles to summarize traffic

3.1.4. Conclusions

HAPviewer can be used to manually check the results of some automated analysis to prove their consistency or to verify bad results. Especially the feature of filtering certain traffic out of the analyzed graph is very useful.

3.2. Beyond Network Telescopes: New Directions based on One-Way Flow Classification [11]

3.2.1. Short overview

This paper discusses a method of monitoring network traffic based on flow data collected within a large network consisting of Cisco switches. The analysis based on flow data is much simpler and efficient compared to using network telescopes. This is because network telescopes require high end equipment and use data collected on a unused subnet, and not data from a live network. This section will only discuss the important parts for our project, which is the technique used to characterize and classify one way flows. If you are interested in what results the work of this paper has lead to, read on at [7].

3.2.2. Used Signs

To classify flow data, a set of signs is introduced. The signs can be very simple as observing the presence of an attribute in one of the packet headers, or as complex as investigating the behavior of an endpoint or end system involved in the flow communication. The signs itself can be present or absent, depending on the rule definition. Following is the list with signs and their explanation as found in the paper.

- **Remote Scanner (TRWscan, TRWnom, HCscan):**
  Flow originates from a host that behaves like a malicious scanner. This signs is divided into 3 categories according to the algorithm used to determine, if this host is a malicious scanner. The Threshold Random Walk (TRW) has two signs, the first identifying a scanner an the second a nominal host.
- **Unused Local Address (GreyIP):**
  Flow is targeted at a local address that is unpopulated.
- **Well-known Service Unreachable (Unreach):**
  Flow is targeted at a local service endpoint known to be a valid service, i.e., having otherwise a significant amount of clients.
- **Retry (Retry):**
  Flow exhibits an average inter-packet arrival time of more than the maximum time-to-live (TTL) value of 255 seconds.
- **Service Sole Reply (Backsc):**
  Only flow seen for remote host during a time interval of 30 minutes targeting a well-known destination port while using an ephemeral source port.
- **Peer-to-Peer (P2P):**
  Flow is targeted at a host tat is inferred to carry P2P traffic.
- **End-Hosts-Communicating (PotOk):**
  Flow involves a host pair that otherwise exchange bidirectional flows.
- **Artifact (Artef):**
  TCP or UDP flow with both port numbers set to zero representing packet fragments without layer 4 header.
- **Bogon (Bogon):**
  Source address belongs to Bogon space.
- **single Packet (Onepkt):**
  Flow comprises one packet only.
- **Large Flow (Large):**
  Flow comprises at least 10 packets or 10240 bytes.
- **Protocol (TCP, UDP, ICMP, OTHER):**
  We label every flow with its IP protocol type for TCP, UDP and ICMP. All other protocols carry the label OTHER.

3.2.3. Rules and classification of one way flows

This section describes the rules used, to match one way flows into different classes. Firstly, we describe the rules with the used signs as shown in table 4. After that, the categories are listed and explained. It is worth noting, that the rules of each specified class do not overlap, so a flow can only be assigned to one of the given classes.

Rule	Used Signs
Malicious Scanning	*{TRWscan, !HCscan, !PotOk}* *{HCscan, !TRWscan, !TRWnom, !PotOk}* *{GreyIP, Onepkt, !TRWscan, !HCscan, !Backsc, !ICMP}* *{TRWnom, HCscan}*
Other Malicious	*{GreyIP, !TRWscan, !HCscan, !Onepkt, !ICMP, !Backsc}* *{Onepkt, !GreyIP, !ICMP, !TRWscan, !HCscan, !Bogon, !P2P, !Unreach, !PotOk, ! Backsc}*
Backscatter	*{Backsc, !TRWscan, !HCscan, !P2P}* *{ICMP, !TRWscan, !HCscan, !TRWnom, !PotOk, !Backsc, !Bogon}* *{GreyIP, TRWnom, ICMP, !HCscan}*
Service unreachable	*{Unreach, !TRWscan, !HCscan, !Bogon}*
Benign P2P Scanning	*{P2P, !GreyIP, !Unreach, !TRWscan, !HCscan, !Bogon}*
Suspected Benign	*{PotOk, !Unreach, !P2P, !TRWnom, !Bogon}* *{Large, !GreyIP, !TRWscan, !HCscan, !P2P, !Unreach, !PotOk, !ICMP, !Backsc, ! Bogon, !TRWnom}* *{TRWnom, !GreyIP, !HCscan, !P2P, !Unreach, !Onepkt, !Bogon}*
Bogon	*{Bogon, !TRWscan, !HCscan, !GreyIP, !Backsc}*

Table 4: List of inspected rules for all classes

The mentioned paper introduces the following classes. Note that the artifact class is being removed because it does not achieve a high coverage because it results of fragmented packets without a header on layer 4. All flows which do not match one of the classes is assigned in to the “OTHER” class. The classes used are described as they are in the paper.

- **Malicious Scanning:**
  Probing for the exploitation of vulnerabilities in end systems.
- **Other malicious:**
  One-way flows that do not match any scan detection rules but are otherwise suspicious.
- **Backscatter:**
  Replies to DoS attack flows using randomly chosen source IP addresses that hide the real identity of the attacker.
- **Service Unreachable:**
  Access attempt to temporary unavailable service.
- **Benign P2P Scanning:**
  P2P applications frequently try to access peers maintained by their local host cache that is not always up-to-date.
- **Suspected Benign:**
  One-way flows may exist as part of benign applications using data and control connections in parallel and employing one of them for acknowledgment only. Another cause may be temporary failures within an otherwise productive communication.
- **Bogon:**
  One-way flow originating from Bogon IP space.

- **Other:**
  One-way flows that do not match any of the above classification rules.

3.3. Evaluating and Improving the Detection of Internet Background Radiation [12]

3.3.1. Short overview

The interesting parts for our project in this article is the applied frequent item set mining (FIM). The used signs and rules to classify all the flows are not mentioned, because this project uses the signs and rules defined by the paper “Beyond Network Telescopes: New Directions based on One-Way Flow Classification” by E. Glatz. The paper mentioned in this section is used to understand the functioning of FIM.

3.3.2. Frequent Item Set Mining (FIM)

In this article, FIM is used for gaining more information about specified classes which make a large percentage out of the total flows. For this purpose, files of the interesting classes where generated, which only contained the information of flows belonging to this class. This reduces the time consumption of the FIM analysis. Applying the FIM analysis, the occurrence of the same port number in various flows can lead to identify malicious traffic, for example a worm or bot trying to connect to its master or a virus scanner checking for updates. Based on packet information, application which use well-known ports to communicate, can be identified.

3.3.3. Conclusions

This paper provides useful information on applying FIM analysis to flow data. It describes some methods used for analyzing the gained results of the frequent item sets. The signs and rules of this article are not important, because this project uses an other rule-set an classification model. But the FIM analysis shows, how the rules which assign flows to specific classes can be considered wright and be improved.

4. Frequent Item Set Mining Theory

4.1. Cisco Netflow Format [13]

The Cisco Netflow format describes IP flows which contain all packets corresponding to a connection. A connection is identified by the 5 Berkeley key elements (source IP, source port, destination IP, destination port and protocol). A new flow entry is created when ever a packet arrives, which has at least one different attribute according to the 5 Berkeley key elements. Flows are gathered to analyze them for security and network device misconfiguration. The attributes saved for every captured flow are listed in table 5.

Attribute	**Description**
Source IP	IP address of sending host.
Destination IP	IP address of receiving host.
Source Port	Port number used on sending host. If ICMP [14] is used, the port number contains the type and code of the ICMP message, computed as follows: type*256+code. Not all routers and switches do set the source port when exporting ICMP messages.
Destination Port	Port number used on receiving host. This field can be zero, for example when receiving ICMP messages.
Next Hop	IP address of the next router on the network path.
Input phys. IF	Index of the physical interface receiving the packet.
Output phys IF	Index of the physical interface sending the packet to the next hop.
Packet count	Number of packets in this flow.
Byte count	Number of bytes transmitted in this flow.
Starting flow time stamp	System uptime of the capturing device when the flow starts.
Ending flow time stamp	System uptime of the capturing device when the flow ends.
IP protocol	Layer 4 protocol type. (ICMP: 1, TCP: 6, UDP: 17)
ToS byte	Type of Service byte as discussed in Chapter 2.2.
TCP Flags	Cumulative OR of all TCP flags. Cisco Netflow Format Version 5 does not set this field.
Source AS	IP address of the AS which sent the packet.
Destination AS	IP address of the AS to which the packet is forwarded.
Source subnet mask	Subnet mask of the sending AS.
Destination subnet mask	Subnet mask of the receiving AS.
Flags	Among other things, this flag indicates, which flows are inactive.

Table 5: Explanation of fields defined by Cisco Netflow format version 5

Flows are collected in the cache of a network device. They are exported to a Netflow collector if the connection is closed or inactive. Because the entries are not in correct timing order, they must be

sorted after exporting them to the Netflow collector. To determine, which flows have expired, the following aging strategies are applied to every flow in the flow table:

- Transport is completed (TCP FIN or RST)
- When the flow cache is getting full, a number of heuristics are applied to age flows in the cache.
- Flows which are idle for 15 seconds are aged to expire.
- Flows which lived over 30 minutes are aged to expire.

The timers for aging flows to expire can be set manually. The idle timer can be configured to expire between 10 and 600 seconds. The timer to define long living flows can be configured to a value between 1 and 60 minutes.

Catalyst Switches use different aging Timers, because they do not have a Netflow cache. They use the MLS (Multi Layer Switching) cache and the following timers:

- Default aging Timer: 256 seconds.
- Fast aging Timer: Disabled
- Long aging Timer: 1920 seconds.

4.2. Explaining the inspected signs

Now we analyze the signs used in the detector rules to identify, which information of the Cisco Netflow information is used to classify the inspected one way flows.

The signs “TRWscan” [15], TRWnom” [15] and “HCscan” [16], for identifying scanning traffic, are not explained, because they base on well-known algorithms to determine malicious scanners. This does not mean, that every scanner is covered in these rules. The analysis of other classes could lead to a new rule, identifying more malicious traffic. Below is a listing explaining the signs defined in table 4.

GreyIP Local IP address which is unused. If during an observation period of 400 hours a local IP address is never the source of communication, but is the destination of traffic, it is considered as a GreyIP

Unreach A valid local service servers at least 20 different clients during an observation period of 400 hours and is defined by the 2-tuple {localIP, local port}. The local port number has to be a well-known port. One way flows targeting a valid local service are labeled as „unreachable“.

Large A flow containing at least 10 packets or 10240 bytes of size. This information is looked up in the packet and byte count fields of the Netflow data structure.

Retry Flows with an average inter-packet arrival time bigger than the maximum TTL (time-to-live) value of 255 seconds.

Backsc If only one flow within 30 minutes is detected from a remote host, targeting a host on a well-known port, it is considered as backscatter if the packet uses a transient source port. Backscatter traffic is caused by DDoS attacks.

P2P Analysis of peer-to-peer communication is done in three different ways, according to researched literature in the work of E. Glatz [11]. The three methods are described below:

- A connection pairing which uses simultaneous UDP and TCP connections not originating or targeting any well-known ports. This analysis is made based on the fact, that known peer to peer protocols use UDP packets for signaling and TCP connections to exchange data.
- The second method is using a list of ports known to be used from peer-to-peer applications. The list contains ports which are normally not used by any other applications.
- Peer-to-Peer applications often negotiate their ports with its peers first and then are propagated to other peers. The port numbers will be above 1024. It has been observed, that peer-to-peer applications don't use parallel connections to negotiate this information.

So the first criteria of this sign are bidirectional UDP communications with port numbers above 1024. Secondly, all hosts with at least 5 peers seen as 5 different IP addresses, having set up a parallel connection with at least one of the peers.

Any flows passing these 3 algorithms within an observation period of 400 hours, are marked as peer-to-peer traffic. Also flows targeting at a peer-to-peer host are marked as peer-to-peer traffic.

PotOk The flow has a host pairing (local- and remote IP) which otherwise have bidirectional and benign communication. Therefore this flow is considered as potentially OK.

Artef Describes a flow that consists of packet fragments without any layer 4 header, for both protocols (TCP and UDP). The source- and destination port of these packets are set to zero. The protocol information is gained from the protocol field of the IP header.

Bogon The IP source address of this flow refers to Bogon address space. Bogon address space is not assigned by any registrars or is not in use at the moment.

Onepkt Defines a signle flow that only consists out of one packet.

TCP, UDP, ICMP, OHTER Refers to the protocol in use. This information is looked up in the protocol field of the IP header. These are 4 different signs.

Knowing what information the signs represent, it is possible to determine what information has to be considered in the frequent item-set mining analysis of the classes chosen to inspect. This is part of the next chapter.

4.2.1. Inspected classes

Within the frequent item set mining described in this chapter, four interesting classes are inspected. These classes show a significant peak in assigned flows compared to other periods of the same class. This seems to lead to lesser assigned flows of the class malicious scanning in the same period. Table 6 lists the classes having a peak in assigned flows.

Class	**Periods**
Other malicious	2008-08
Backscatter	2007-02, 2007-08, 2008-02
Unreachable	2008-08, 2009-02
Benign P2P	2007-08, 2008-02

Table 6: Inspected classes and their according periods

If there is enough time, a frequent item set analysis is considered over the class “Other”. This analysis is considered for trying to reduce the percentage of one way flows not matching any of the rules specified.

4.3. Preparations

4.3.1. How to analyze the given information

Flows are preprocessed because they are not in correct timing order, when they are gathered by a netflow collector. The preprocessed flows are stored in files compressed with the gzip algorithm and therefore have the extension “*.gz”. Corresponding to every file containing flows, a file with the according sign sets exists. The sign sets are listed in the same order as the flows are. The ending of the sign set files is “*.sig.gz” and are also compressed with the gzip algorithm. Each of the flow files

contains all the flows of an interval during 10 minutes. The flow and sign files are named with the date and time the flows when the flows where. Following is an example on how the files are named:

EXAMPLE: I_YYYYMMDD.hhmmss.gz

- YYYY = year (2004..2010)
- MM = month of year (01..12)
- DD = day of month (01..31)
- hh = hour of day (24 hours per day, 0..23)
- mm = minutes of hour (0..59)
- ss = seconds of minute (0..59)

For analyzing the given data with the sam tool described in the next section, the flows must be converted into a text file. The Conversion is done by extending the sign_eval2 tool as described in the next section „4.3.2 Software for FIM Preparation“.

To gain information about the most common attributes of the IP flows inspected, the FIM analysis inspects the data and finds out the maximum item-sets with a support greater than 10%. The 10% are chosen because an attribute which appears in more of 10% of the cases is considered as a large enough portion of the data to identify a special behavior of the flows in the analyzed class. If we see, that the analysis with this support is inappropriate, the support value is redefined.

Analyzing all the flows for a given period of 400 hours is quite a lot of data. To process all files corresponding to a period consumes a lot of time and resources. Also the output files would consume up to 370GB space. Therefore, we first choose a random interval of three continuous hours per peak period and analyze the contained flows and item-sets in this three hour interval.

When inspecting only one interval of three hours from a time period, it could be that the chosen interval does not cover the cause of the peak showed in the statistics made by E. Glatz. This is especially the case, when the traffic, which caused the peak, is bursty or in other ways limited to a time window. This behavior could be analyzed in a long term analysis over a whole period of 400 continuous hours. Performing a long term analysis has another benefit. It shows, if the chosen interval is a good over all representation of the whole period.

4.3.2. Software for FIM Preparation

For analyzing the right flows, E. Glatz has written the software sign_eval2 [18] stored in the software folder of the project documentation root, which reads a flow file, the corresponding sign file and a rule file. It then checks, that the lines in the sign file correspond to the line in the flow file. After this check, the tool checks, if the signs of a flow match one of the rules in the specified rule file. If the signs match a rule, this is recorded for statistical purposes.

To only analyze flows and signs of an inspected class, the sign_eval2 tool is extended to support the writing of flows and their signs applying a rule-set of a given class. Flows and their signs are stored in two separate files in the same directory. Output files are formated in „*.csv“ style, having a header line for each 10 minute sample. The extended sign_eval2 (sign_eval2_for_FIM) tool does append information to the output files, so it is possible to write output files containing more than only a 10 minute interval of the flows, by passing the same output filename to the sign_eval2_for_FIM tool. The filename of the output sign file is derived from the name of the flow output file, so only one output filename has to be passed to the extended tool.

4.3.3. Frequent Item-Set Mining software

For applying the frequent item set analysis to the gathered sign combinations on corresponding flows, the software sam [17] is being used. This software uses a split and merge algorithm to verify the item sets. To start the analysis, the following command is entered in a command line:

```
# ./sam -s$support [-f “$separator”] -tm $inputfile $outputfile
```

- -s$support
  This option specifies the support in percentage of the item sets being written to the output file. $support is replaced with the percentage.

- [-f “$separator”]
  The normal item separator is a space, but if the input file uses an other separator, it is possible to specify it whit the -f option.
- -tm
  This specifies, which target type of item sets are being analyzed. The “m” attribute specifies that the maximal item sets should be listed.
- $inputfile
  A text file containing the items to analyze. The default separator for items is a space. To separate transactions, the special character \n (newline) is used.
- $outputfile
  All results of the frequent item set mining analysis will be stored in this file

4.3.4. Performance

The Performance of an analysis is very important due to limited resources. This section will give an overview over memory consumption and estimated computing times.

The sign_fltest2 tool does not need much memory for analyzing up to 3 million flows. Flows are represented by cflow structs, which need 56 byte. Once initialized, the cflow structs are handled with pointers, which ensures, that a flow is only allocated once. The memory consumption for 3 million flows then would be about 160 Mega byte, assuming that 1 Mega byte has 1024 Kilobyte.

Sam does not need a lot of memory, because it uses an array as it's only data structure as described on the web page of the developer.

Analyzing 2 million flows per 10 minute file is quite a lot of work. Because the sign_eval2 tool first checks if the signs and flows are in correct order, the whole data structure has to be traversed twice. The second traversation is for matching the rules specified in the rule file against each sign set of a flow. If the sign set matches the rule, the flow and it's according sign set are written to the output files. Therefore, the sign_eval2 tool takes up to 1,5 hours for writing the matching flows and their signs to the output files.

Over a interval of three continuous hours of a period, the sam tool is very fast. It only takes one minute at maximum for applying the FIM analysis to the input data. This is the main reason, why the analysis of all 10 minute flow files is only done once for proving that the item-sets over a 10 minute flow file correspond to the item-sets produced by the FIM analysis over the whole 3 hours.

5. Frequent Item-Set Mining Analysis

5.1. Inspected intervals

In the analysis, we inspect the before mentioned classes in table 6 over the most interesting periods, showing a peak in assigned flows. Table 7 contains the source directories, where the input files of to the according time period can be found. Also shown in this table are the inspected intervals of the corresponding periods. Note that flows and their according signs are stored in two separate files.

Period	Time Interval
2007-02	28th January 2007 14.00 to 17.00
2007-08	31th July 2007 00.00 to 03.00
2008-02	03th February 2008 02.00 to 05.00
2008-08	31th July 2008, 03.00 to 06.00
2009-02	30th January 2009 07.00 to 10.00

Table 7: Showing Time periods and the inspected interval

Each directory listed in table 7 contains flow files over a period of 400 continuous hours. The table lists also time and date of the inspected interval. The three hours interval is covered in 18 continuous flow and sign files. Three hours are chosen, because the resources and the amount of time are not that big for analyzing such a small interval. Spending lesser time on the analysis leads to faster results when interpreting the item-sets of the FIM analysis. Only inspecting three hours can lead to missing the cause of a peak period or to not investigating important item-sets. This can be corrected by making a FIM analysis over the whole period of 400 continuous working hours.

5.2. Expectations

The main task acquired with the FIM analysis is to evaluate the specified classes and their assigned rules. The analyzed data helps to identify, if the rule-set classifies a high percentage of the inspected flows correctly. This is done by explaining the sign item-sets. This does not mean, that every single flow can be identified and classified correctly, afterwards.

Having a look at the gathered flow item-sets is the second target. Explaining the possible causes of the flows assigned to an item-set verifies, if the flows matching the item-set belong to the class they have been assigned to.

In the paper [11] written by E. Glatz and Xenofontas Dimitropoulos, peaks can be observed in various periods of the classes „Other Malicious“, „Backscatter“, „Unreachable“ and „Benign P2P“. The third task acquired with the FIM analysis is to find the cause of the peak from a given class. Traffic causing a peak in a class does not always have to be present, for example, when the traffic is very bursty and occurs only on one or two days of the inspected period. If this is the case, we might not cover the cause of the peak with an analysis over a random 3 hours interval and therefore an over all analysis, covering the whole 400 hours of the peak periods, must be made. An over all analysis is made by splitting the 400 hours into 3 hour intervals. The 3 hour intervals of flows and signs are then analyzed separately with the FIM tool `sam`.

We expect that an over all analysis leads to the cause of the peak in a period. Another expectation is that the over all analysis will show, that the random three hours intervals taken for our analysis are a good representation for the whole 400 hour period. The peak interval is identified by comparing all three hour intervals of a period against each other.

When a pattern is found in the inspected flows showing that the flows are assigned to the wrong class, the rules of the class they belong to will be expanded, so that the flows can be assigned correctly.

Recording statistics over the occurrence of signs of a inspected interval is the last target. These statistics allows us to verify, which rule assigns the most flows to the class. In these terms, we do also

record the occurring signs for all flow item-sets to see, if the flows of the item-set are assigned by the same rule or by different rules.

5.3. Analysis Setup

First an analysis was made, which wrote flow and sign output files covering a 10 minute interval according to its input file. made files according to each interval. Then the observation was made, that the analysis of a continuous interval of three hours only takes one minute in maximum. Therefore a script is written, to merge all flows of a 10 minute sample file into one file covering the whole continuous three hour interval. The scripts are appended to the appendix.

The execution of the scripts to write the output flow and sign files have been started with the nohup command, which allows a user to logout after he started the job, because nohup sets the parent process to init() and not to the calling parent shell. The nohup command itself prints all output to a file, the *nohup.out* file. Therefore it is wise to redirect all output to a file specified by the user.

The information, where the gained result files are stored on the Scylla cluster of the ETH are listed in the appendix.

5.3.1. Style of referencing item-sets

For having references to item-sets listed in the tables, a reference scheme is introduced as described in table 8.

Style	Explanation
F$	If the item-sets are not listed in the table, the flow item-sets are referenced with the capital letter F . The $ sign is substituted with a number which is unique for a flow item-set within the whole section.
S$	When not listing the item-sets in the table, sign item-sets are referenced with the capital letter S. The $ sign is substituted with a number which is unique for a sign item-set within the whole section.
#$	When referencing an item-set within the conclusions, they will be marked with a leading route character (#) and a number, represented by the $ sign. The number is uniquely identifiable over the whole document. So it is possible to refer to an item-set mentioned in a previous section or chapter of this document.

Table 8: Flow and sign item-set reference scheme

5.4. Results Class Other Malicious (Support 10%)

In this interval we see the following maximum frequent item sets with the defined support of 10%. Note that the following Items are not listed in table 8: *ASremote:4134*, *ASlocal:559* and *flowtype:2*. The items are not listed, because they appear in all flow item-sets including the one over the whole 3 hour interval. Shrinking the item-sets has the benefit of displaying them in one line without break. The following item-sets have been reported. The value in brackets refers to the heading of table 9.

- *LIP:129.132.2.21 Lport:37 tos:0 dOctets:46 dPkts:1 durMs:0 prot:UDP* (F1)
- **{#1}** *GreyIP* (S1)
- *Onepkt UDP* (S2)

File	Flow Item-set (F1)	Sign item-set 1 (S1)	Sign item-set 2 (S2)
I_20080731.030000	68%	12.2%	85.1%
I_20080731.031000	68.1%	12.3%	85.7%
I_20080731.032000	68.2%	12.1%	84.2%
I_20080731.033000	69.7%	12.6%	86.5%
I_20080731.034000	68.8%	12%	87.2%
I_20080731.035000	69.4%	12.2%	87%
I_20080731.040000	65.7%	12.2%	84.1%
I_20080731.041000	68.8%	11.7%	87%
I_20080731.042000	68.6%	12.2%	86.2%
I_20080731.043000	70.2%	11.8%	87.4%
I_20080731.044000	70.3%	11.4%	86.8%
I_20080731.045000	70.6%	10.3%	86.9%
I_20080731.050000	72.9%	11.1%	88.3%
I_20080731.051000	73.6%	11.2%	88.2%
I_20080731.052000	74.6%	10.2%	89%
I_20080731.053000	75.9%	-	89.7%
I_20080731.054000	75.7%	-	89.9%
I_20080731.055000	75%	-	88.4%

Table 9: Percentage of flow and sign item-sets over inspected 10 minutes intervals

Furthermore, an overall analysis is made with two files, one containing all inspected flow data, the other containing all signs according to the inspected flows. For the ability to display the flow item-set in a single line, the items *flowtype:2*, *ASremote:4134* and *ASlocal:559* are not listed. Table 10 lists the item-sets found in the analysis over the whole three hours.

File	Item-sets (flows and signs)	Pct
merged_flows_aug2008	LIP:129.132.2.21 lPort:37 tos:0 dOctets:46 dPkts:1 durMs:0 prot:UDP	71.10%
merged_signs_aug2008	**#1** GreyIP **#2** OnePkt UDP	11.2% 87.3%

Table 10: Flow and sign item-sets over the whole 3 hour interval

5.4.1. Conclusions

Table 9 shows the item-sets found in all 10 minute flows, each file analyzed for itself. This was made to check, if the item-sets found correspond to the analysis made over the whole interval of three continuous hours. The observed flow item-sets are all the same. We only note a change int the occurrence of sign item-sets. The last three 10 minute intervals don't show the *GreyIP* {**#1**} item-set. This does not mean, that it does not exist, it simply has a to small percentage of occurrence and therefore slides below the border of a support of 10%. Table 10 lists the item-sets found by analyzing the files covering the whole interval at once.

The results show, that the most traffic classified as „Other Malicious“ traffic is caused by clients who send Time Protocol requests. The IP 129.132.2.21 in the results corresponds to the server swisstime.ethz.ch. Konstantinos Karampogias has also come to the same solution in his paper [12]. The cause of this large amount could be misconfiguration of clients. What is stunning about these facts is, that the UDP datagrams sent to the server, are above the minimum packet length of 28 bytes (20 bytes IP header, 8 bytes UDP header).

Another exploration is that 71% of the flows classified as „Other Malicious“ are routed from the source AS 4143 [19]. Further investigations show that a variant of the sober worm [20] uses a time stamp to establish its date. In the article mentioned, the timeserver swisstime.ethz.ch is not mentioned, but nevertheless it could be possible, that the worm is using this server in newer variants, because the information from Avira is a bit old. The worm needs the time stamp to estimate, if the current date is past march the 24th of 2004. If this is true, the worm downloads an additional file from a web server. The overview of activities on this port can be viewed by SANS [21]. All left to do, is to adjust the timing window. The link to SANS shows, that a large amount of traffic is directed to this port even though the time protocol is obsolete. Another explanation for these findings is that a lot of clients try to use a time server which only serves as NTP server and does not offer the obsolete time Protocol.

Interpreting the signs found in the maximum item-set leads to the conclusion, that 87.3% of all flows in this class are transmitted with the UDP protocol. Covering this amount of flows with the *Onepkt* sign set, indicates a malicious activity, because benign flows normally have more than one packet assigned to them. It can not be assumed, that the *GreyIP* item-set {**#1**} covers a group of flows which are not assigned to the other item-set {**#2**} found, because the *GreyIP* sign can also be set on flows covered by the other item-set.

5.5. Class Other Malicious (Support 5%)

Table 11 does not list the following items in the flow item-sets: *ASremote: 4143*, *ASlocal:559* and *flowtype:2*.

File	Item-sets	Pct
Flows	prot:TCP tos:0	7.3 %
	#3 IIP:129.132.2.21 IPort:37 tos:0 dOctets:46 dPkts:1 durMs:0 prot:UDP	71.1%
Signs	**#4** TCP GreyIP	5.8%
	#6 TRWnom Onepkt	5.8%
	#7 TRWnom UDP	5.0%
	#5 GreyIP UDP	5.4%
	Onepkt UDP	87.3%

Table 11: Item-sets of class Other Malicious with support of 5%

5.5.1. Conclusions

What we can observe is that the two main Item-sets, already listed in table 10, have the same percentage. The small flow item-set listed in table 11 is not very interesting, because it does not reveal any information about the causes of the flows.

Interesting is, that the item-set {**#3**} has the same percentage as with a support of 10%. Not having that item-set split up in finer grained item-sets leads to the conclusion, that the main traffic in this class refers to this item-set.

The item-sets of the signs show that the *GreyIP* item-set is now splitted up into two item-sets: *TCP GreyIP* {**#4**} and *GreyIP UDP* {**#5**}. Counted together the same percentage results as the single signed *GreyIP* item-set listed in table 10. This illustrates that the TCP and UDP protocol are equally popular among this class.

The analysis has also brought up two new sign item-sets ({**#6**}, {**#7**}), both having the *TRWnom* sign set. These two item-sets could indicate malicious scanning traffic, because the *TRWnom* sign does cover slow scanning types having a long gap between two sent packets. Further inspections are needed to verify if it is indeed scanning traffic and therefore should be matched to the „malicious scanning“ class.

5.6. Results Class Backscatter (Support 10%)

Tables 12, 13 and 14 hold the discovered item-sets covering the intervals of each inspected period. The item-set do not display the *flowtype:2* and *ASlocal:559* information, because they appear in every item-set.

File	Item-sets	Pct
flows_feb 2007	**#9** rPort:11 lPort:0 prot:ICMP	12.5%
	#8 dOctets:244 dPkts:4 tos:0 rPort:8 lPort:0 prot:ICMP	12.4%
	#10 tos:128 lPort:0 prot:ICMP	14.6%
	dPkts:1 durMs:0 tos:0 lPort:0 prot:ICMP	15%
	#8 dPkts:1 durMs:0 rPort:8 lPort:0 prot:ICMP	13.3%
	#8 dOctets:122 dPkts:2 tos:0 rPort:8 lPort:0 prot:ICMP	34.6%
signs_feb 2007	Onepkt GreyIP ICMP	10.30%

Table 12: Resulting item-sets of inspected interval in February 2007

File	Item-sets	Pct
flows_aug2007	**#8**, **#10** Tos:128 rPort:8 prot:ICMP	10.9%
	#8 dOctets:244 dPkts:4 tos:0 rPort:8 prot:ICMP	13.4%
	#8 dPkts:1 durMs:0 rPort:8 prot:ICMP	10.6%
	dPkts:1 durMs:0 tos:0 prot:ICMP	14.2%
	#8 dOctets:122 dPkts:2 rPort:8 tos:0 prot:ICMP	39.6%
signs_aug2007	Onepkt ICMP	18.8%
	GreyIP ICMP	48.6%

Table 13: Resulting item-sets of inspected interval in august 2007

File	Item-sets	Pct
flows_feb2008	dOctets:112 lPort:0 prot:ICMP	10.2%
	IIP:192.33.90.66 lPort:0 prot:ICMP	11.9%
	#8 dOctets:244 dPkts:4 rPort:8 tos:0 lPort:0 prot:ICMP	12.1%
	#9 dOctets:56 dPkts:1 rPort:11 durMs:0 lPort:0 prot:ICMP	13.9%
	#10 tos:128 lPort:0 prot:ICMP	16.1%
	#8 dOctets:122 dPkts:2 rPort:8 tos:0 lPort:0 prot:ICMP	17.8%
	dPkts:1 durMs:0 tos:0 lPort:0 prot:ICMP	16.7%
	#8 rPort:11 dPkts:2 lPort:0 prot:ICMP	10%
	#8 rPort:11 tos:0 lPort:0 prot:ICMP	23.1%
signs_feb2008	GreyIP ICMP	24.7%
	Onepkt ICMP	28%

Table 14: Resulting item-sets of inspected interval in February 2008

5.6.1. Conclusions

The results in table 12, 13 and 14 assume that all frequent item-sets are using the ICMP protocol. If we take a closer look at some of the item-sets, we find that the remote port number is not 0 according to the specification of ICMP, which has no port numbers defined. This is because the Cisco Netflow format will encode the ICMP type and code as the remote port. Some of our item-sets have their remote port set to 8 {**#8**} or 11 {**#9**}. The ICMP types are represented as a decimal value, therefore these types do not exist. Having wrong ICMP types and codes set as values for *rPort* can be caused by a missconfigurated bit mask on a router as described in this newsgroup article [22].

Another interesting observation is the IP address 192.33.90.66 which belongs to a host of the planet flow network, maintained by the ETH Zurich. The node is part of a research platform to determine new network services as described in [23]. On the monitoring page of the project it says, that the nodes send each other trace route commands, which would explain some of the echo requests observed in the frequent item-sets. This fact needs some further investigation over a longer period of time to determine, if a new rule must be created, or the existing rules lack of some signs, which must have been set to classify the flows correctly. To match these flows to the right class the IP address 192.33.90.66 could be recorded in the list, which contains benign hosts with their according services.

Another item in the item-sets was the precedence of datagrams having their ToS byte set to 128 {**#10**}. This indicates a network precedence having the most significant bit set. As defined in RFC 791 [24], the priority is in the middle of the range. If a network is under heavy load, the use of a precedence forces the network to prefer packets upon the set precedence. Applying the control mechanism is up to each network and is intended to use by gateway control originators only. An attacker could set the ToS byte accordingly, to ensure that the packets of his attack are preferred, if the victim network implements this congestion control. Further investigations have shown, that some ISP's, for Example China Telecom, do implement this mechanism in their networks. We do not list all ISP's or company networks implementing this mechanism, because the effort of assigning 173'949 IP's to their range is really time consuming.

All other flow item-sets do not give much information about the classified flows. They just show, that a high percentage of the one way flows matched to the class „Backscatter“ use the ICMP protocol.

When looking at the frequent item-sets of the signs, it is a bit confusing, that no frequent item-set with the sign Backsc is found. This means, that there are not enough flows with this sign, or that the sign is combined with many other possible signs and therefore the support used in the FIM analysis is to high to identify them.

Analyzing the sign item-sets of this class does not reveal enough information to say, if the rules cover a high percentage. A statistical overview of signs over all flows assigned to this class is needed to make a statement, which rule is the most effective in this class.

5.7. Class Backscatter (Support 5%)

Tables 15, 16 and 17 hold the discovered item-sets covering the intervals of each inspected period. The item-set do not display the *flowtype:2* and *ASlocal:559* information, because they appear in every item-set.

File	Item-sets	Pct
Flows feb2007	**#19** durMs:2368 dOctets:122 rPort:8 dPkts:2 lPort:0 prot:ICMP	5.2%
	#19 durMs:2240 dOctets:122 rPort:8 dPkts:2 lPort:0 prot:ICMP	5.2%
	#11, #19 ASremote:9318 tos:0 rPort:8 lPort:0 prot:ICMP	5.2%
	#19 durMs:2304 dOctets:122 rPort:8 dPkts:2 lPort:0 prot:ICMP	5.5%
	dOctets:112 dPkts:2 lPort:0 prot:ICMP	5.7%
	#14 rPort:771 lPort:0 prot:ICMP	6.8%
	#18, #19 tos:160 rPort:8 lPort:0 prot:ICMP	6.8%
	dOctets:56 dPkts:1 lPort:0 durMs:0 prot:ICMP	7.4%
	#19 dOctets:61 tos:0 rPort:8 dPkts:1 durMs:0 lPort:0 prot:ICMP	6.4%
	#20 rPort:11 durMs:0 lPort:0 prot:ICMP	5.3%
	#20 rPort:11 tos:0 lPort:0 prot:ICMP	7.6%
	#19 dOctets:244 dPkts:4 tos:0 rPort:8 lPort:0 prot:ICMP	12.4%
	#21 tos:128 durMs:0 lPort:0 prot:ICMP	5.1%
	#19, #21 tos:128 dOctets:122 dPkts:2 rPort:8 lPort:0 prot:ICMP	5.6%
	#19 dOctets:122 dPkts:2 tos:0 rPort:8 lPort:0 prot:ICMP	34.6%
Signs feb2007	**#23** TRWnom GreyIP ICMP	6.0%
	#22 Onepkt GreyIP ICMP	10.3%

Table 15: Resulting item-sets of class Backscatter with 5% support in February 2007

File	Item-sets	Pct
Flows aug2007	**#12, #19** ASremote:7738 rPort:8 tos:0 prot:ICMP	5.4%
	#20 rPort:11 prot:ICMP	5.9%
	#19 durMs:2432 dOctets:122 dPkts:2 rPort:8 prot:ICMP	5.9%
	#19 durMs:2176 dOctets:122 rPort:8 dPkts:2 prot:ICMP	5.9%
	#19 durMs:2240 dOctets:122 tos:0 dPkts:2 rPort:8 prot:ICMP	5.3%
	#19 durMs:2304 dOctets:122 rPort:8 tos:0 dPkts:2 prot:ICMP	5.5%
	#19 durMs:2368 dOctets:122 rPort:8 tos:0 dPkts:2 prot:ICMP	5.6%
	#13, #19 ASremote:4766 dOctets:122 rPort:8 dPkts:2 tos:0 prot:ICMP	7.0%
	#19 dOctets:61 rPort:8 tos:0 dPkts:1 durMs:0 prot:ICMP	6.5%
	#14 rPort:771 dPkts:1 durMs:0 prot:ICMP	5.3%
	#14 rPort:771 tos:0 prot:ICMP	7.0%
	#19, #21 tos:128 dOctets:122 dPkts:2 rPort:8 prot:ICMP	8.0%
	#19 dOctets:244 dPkts:4 tos:0 rPort:8 prot:ICMP	13.4%
Signs aug2007	**#24** Onepkt GreyIP ICMP	9.20%

Table 16: Resulting item-sets of class Backscatter with 5% support in august 2007

File	Item-sets	Pct
Flows feb2008	**#15** rPort:0 lPort:0 prot:ICMP	5.5%
	#16, #17 lIP:129.132.2.21 rPort:8 tos:0 lPort:0 prot:ICMP	5.5%
	#13, #19 ASremote:4766 rPort:8 tos:0 lPort:0 prot:ICMP	6.3%
	#20 dOctets:112 rPort:11 dPkts:2 tos:0 lPort:0 prot:ICMP	7.0%
	#17, #20 lIP:192.33.90.66 dPkts:1 rPort:11 durMs:0 lPort:0 prot:ICMP	5.1%
	#17 lIP:192.33.90.66 dPkts:1 tos:0 durMs:0 lPort:0 prot:ICMP	6.0%
	#17, #20 lIP:192.33.90.66 rPort:11 tos:0 lPort:0 prot:ICMP	5.7%
	#19 dOctets:244 dPkts:4 rPort:8 tos:0 lPort:0 prot:ICMP	12.1%
	#20 dOctets:56 dPkts:1 rPort:11 tos:0 durMs:0 lPort:0 prot:ICMP	8.2%
	#20 tos:128 dPkts:1 durMs:0 lPort:0 prot:ICMP	7.6%
	#20, #21 tos:128 rPort:11 lPort:0 prot:ICMP	5.6%
	#21 tos:128 dPkts:2 lPort:0 prot:ICMP	6.2%
	#19, #21 tos:128 rPort:8 lPort:0 prot:ICMP	8.9%
	#19 dOctets:122 dPkts:2 rPort:8 tos:0 lPort:0 prot:ICMP	17.8%
	#19 dPkts:1 durMs:0 rPort:8 lPort:0 prot:ICMP	6.8%
Signs feb2008	**#25** Large ICMP	10.0%
	#22 GreyIP ICMP	24.7%
	#22 Onepkt ICMP	28.0%

Table 17: Resulting item-sets of class Backscatter with 5% support in February 2008

5.7.1. Conclusions

Setting the support for the FIM analysis to 5 percent results in finer grained item-sets. With a smaller support, item-sets with only one different item can be observed, which allows a closer look on the inspected flows. Following, only the most interesting item-sets listed in tables 15, 16 and 17 are discussed.

The analysis shows that a lot of false ICMP type and code messages are sourced from three different AS, AS 9318 {**#11**} listed in table 14, AS 7738 {**#12**} and AS 4766 {**#13**} in table 15. The problems of having false ICMP type and codes can be caused by a misconfigured router as described in [22].

The ICMP flows having the remote port set to 771 {**#14**} are messages of type 03 with code 03 which means port unreachable. Port unreachable messages can result from a malicious scanner trying to identify open ports from a host or a range of hosts.

Another type of ICMP flows {**#15**} can be observed in the listing of table 16. This type has the *rPort* set to 0, which indicates, that this router does not record the ICMP type and code or it is a reply of an echo request having the type and code set both to 0.

Item-sets having local IP's as their items correspond to preview findings in the analysis of this class with a support of 10%. IP 129.32.2.21 {**#16**} corresponds to the time server swisstime.ethz.ch and 192.133.90.66 {**#17**}corresponds to the planet flow network.

Another observed fact is that there is a small amount of ICMP echo requests with the ToS field set to 160 {**#18**}. This indicates CRITIC/ECP traffic. Some of the flows do have a ToS value of 128 {**#21**}, which is a level lower than with a ToS value of 160.

A lot of flow item-sets have the remote port set to 8 {**#19**} or 11 {**#20**}. These ICMP code and type values are not defined. Therefore these values could result from a misconfiguration of the bit mask used on the routers generating these flows.

When looking at the sign item-sets, the marked {**#22**} do also appear in the FIM analysis with a support of 10 % and are listed with the same percentage. A new item-set {**#23**} can be found in table 14. It has the item *TRWnom* which indicates, that the analyzed one way flows could be caused by malicious scanning activities. {**#24**} is a combined item-set of the ones gathered in the analysis in august 2007 with a support of 10%. The inspected three hour interval of February 2008 shows a new item-set {**#25**} with an occurrence of 10%. It was not showed in the analysis with a support of 10% because only item-sets with a occurrence of more than 10% are listed. The item-set show that the Large sign along with the ICMP sign is set. Large means, that the flow has at least 10 packets or 10240 bytes of in length.

Applying the analysis with a smaller support does not reveal any information about the covered percentage of flows in this class.

Also surprising in the analysis with a smaller support is, that no item-sets were found having the the *Backsc* sign set. They might not be listed, because the item-sets having the *Backsc* sign set are too small for a support of 5%. This brings up the conclusion, that a small amount of flows are matched to this class, having the *Backsc* sign set.

5.8. Results Class Benign P2P (Support 10%)

Tables 18 and 19 show the resulting item-sets of the class Benign P2P with a support of 10%. The items *ASlocal:559* and f*lowtype:2* are not shown, because they appear in every item-set. The item *flowtype* is only shown, when it has a value differing from 2.

File	Item-sets	Pct
fim_flows_aug2007	**#26** rPort:31415 tos:0 prot:UDP	11.4%
	#31 dPkts:3	10.4%
	#30 tos:128 prot:UDP	11.2%
	#31 prot:TCP tos:0	14.1%
	#31 dPkts:2 prot:UDP tos:0	19.1%
	#31 dPkts:1 durMs:0 prot:UDP tos:0	21.4%
fim_signs_aug2007	TCP P2P	20.3%
	#32 PotOk TRWnom UDP P2P	13.5%
	#33 TRWnom Onepkt UDP P2P	10.4%

Table 18: Resulting item-sets of class Benign P2P with 10% support in august 2007

File	Item-sets	Pct
fim_flows _feb2008	**#29** flowtype:10 prot:UDP	10.6%
	#31 dPkts:2 tos:0	10.1%
	#31 dPkts:2 prot:UDP	10.4%
	#30 tos:128 prot:UDP	11%
	#27 dOctets:46 IIP:82.130.102.218 tos:0 prot:UDP dPkts:1 durMs:0	10.2%
	#27 dOctets:46 durMs:0 IIP:82.130.102.161 tos:0 prot:UDP dPkts:1	12.6%
	#28 dOctets46 durMs:0 IPort:4246 tos:0 dPkts:1 prot:UDP	15%
	#27, #28 IIP:82.130.102.218 IPort:4246 tos:0 prot:UDP	25%
	#27, #28 dPkts:1 durMs:0 IIP:82.130.102.161 IPort:4246 prot:UDP	10%
	#27, #28 IIP:82.130.102.161 IPort:4246 tos:0 prot:UDP	31%
fim_signs _feb2008	PotOk UDP P2P	10.6%
	Large UDP P2P	20.4%
	Onepkt UDP P2P	37.7%

Table 19: Resulting item-sets of class Benign P2P with 10% support in February 2008

5.8.1. Conclusions

Tables 18 and 19 hold some interesting results. The observed remote port 31415 on UDP protocol {**#26**} is used by the XBSlink [25] application which defines a proxy link system for Xbox 360 and PS2/3. The proxy can connect to any other user over the Internet, so that the console, on which the game is played, thinks that its opponent or the cloud playing the same game are on the same local network.

The IP's 82.130.102.161 and 82.130.102.218 {**#27**} are fake e-mule servers, also discovered in the paper by K. Karampogias [12]. This leads to the conclusion, that the UDP port 4246 {**#28**} is also associated to the e-mule service.

Another observation is the flowtype 10 {**#29**}, which is a combined flow type of type 2 and 8. Flowtype 2 describes an inflows, whereas the type 8 defines a unibiflow, meaning that a uniflow between the involved hosts exists, that have otherwise bidirectional communication. Therefore the type 10 can represent failing connections between a host pairing which also have bidirectional connections between them. Analyzing these datagrams can lead to two conclusions. First, the flow is classified correctly, because UDP datagrams can be used to send control messages to a given application. Secondly it could indicate connection failures and therefore would be classified wrong, belonging to the class „unreachable“.

We also notice flows with the ToS byte set to 128 {**#30**}. This could be a hint to a application setting this byte accordingly to improve it's service on congested networks, but it is still up to the network to implement this control mechanism. Another possibility is an ISP having implemented this mechanism.

All other flow item-sets {**#31**} are not very meaningful, because they have items, which can apply for benign or malicious flows. Without any other attributes, it is not possible to identify the cause of the one way flows having these item-sets.

Looking at the sign item-sets, we notice that all of the item-sets have the sign *P2P* in them. If this sign is set, the one way flows normally are originating from a source port normally assigned as a standard port for peer 2 peer applications, such as a torrent client.

Therefore the only interesting sign item-sets are the ones having the *TRWnom* sign set. One item-set {**#32**} has the *PotOk* sign set, which indicates, that the one way flow between this host pairing has otherwise bidirectional connections established. A cause for one way flows for this sign could be a peer, checking if the other peer is still alive, or to send control information to the other peer. When having peer 2 peer traffic, UDP datagrams can also be used, to inform the other peer over new parts of files which ready to seed.

The other item-set {**#33**} has the *Onepkt* sign set, which can be caused by a peer scanning another peer to determine its service port of the P2P application. A malicious cause for these UDP datagrams can be a passively scanning P2P worms looking for unpatched versions of any peer to peer applications.

The resulting flow item-sets listed in table 18 do all have the *P2P* sign set. The item-sets cover almost 70% of the assigned one way flows, which is a large amount. But it can not be said, that the rest of flows do not have the *P2P* sign set only because they are not shown. The reason is that item-sets with a support smaller than 10% are not shown, but may exist.

5.9. Class Benign P2P (Support 5%)

Tables 20 and 21 show the resulting item-sets of the class Benign P2P with a support of 5%. The items *ASlocal:559* and *flowtype:2* are not shown, because they appear in every item-set. The item flowtype is only shown, when it has a value differing from 2.

File	Item-sets	Pct
Flows aug2007	**#34** lPort:23753 lIP:128.178.176.86 prot:UDP	5.0%
	dOctets:120 dPkts:2 prot:UDP tos:0	5.2%
	dOctets:182 dPkts:2 tos:0 prot:UDP	5.0%
	dOctets:144 dPkts:3 prot:TCP	5.1%
	#41 lPort:17543 prot:UDP lIP:134.21.2.59	5.7%
	dPkts:6 tos:0	5.7%
	dOctets:91 tos:0 dPkts:1 durMs:0 prot:UDP	5.6%
	#36 dOctets:60 rPort:31415 flowtype:10 tos:0 dPkts:1 durMs:0 prot:UDP	6.9%
	dPkts:3 prot:TCP tos:0	5.4%
	dPkts:3 tos:0	7.8%
	tos:128 dPkts:1 prot:UDP durMs:0ort:4254 dPkts:1 lIP:82.130.102.161 durMs:0 prot:UDP	7.8%
	flowtype:10 dPkts:2 tos:0	5.3%
	prot:TCP tos:0	14.1%
Signs aug2007	TCP P2P	20.3%
	PotOk TRWnom Onepkt UDP P2P	7.4%

Table 20: Resulting item-sets of class Benign P2P with 5% support in august 2007

File	Item-sets	Pct
Flows feb2008	**#37** dOctets:1610 dPkts:35 lPort:4246 prot:UDP	5.3%
	#35, #37 ASremote:3269 lPort:4246 tos:0 prot:UDP	5.1%
	#35 dOctets:92 dPkts:2 lPort:4246 tos:0 prot:UDP	5.3%
	#35, #37 ASremote:3352 lPort:4246 tos:0 prot:UDP	5.9%
	dPkts:3 tos:0 prot:UDP	6.0%
	#39, #40 lPort:4254 flowtype:10 dOctets:46 tos:0 dPkts:1 durMs:0 prot:UDP	6.7%
	#39 lPort:4254 dPkts:1 lIP:82.130.102.161 durMs:0 prot:UDP	5.5%
	tos:128 dPkts:1 durMs:0 prot:UDP	5.6%
	#37 tos:128 lPort:4246 prot:UDP	7.4%
	#37, #38 dOctets:46 lIP:82.130.102.218 lPort:4246 tos:0 dPkts:1 durMs:0 prot:UDP	6.8%
	#37, #38 dOctets:46 durMs:0 lIP:82.130.102.161 lPort:4246 tos:0 dPkts:1 prot:UDP	8.2%
Signs feb2008	PotOk Onepkt UDP P2P	9.3%
	Large UDP P2P	20.4 %

Table 21: Resulting item-sets of class Benign P2P with 5% support in February 2008

5.9.1. Conclusions

The resulting item-sets listed in tables 20 and 21 show item-sets providing a lot of information. A ripe lookup [26] assigns the IP address 128.178.176.86 of the item-set {**#34**} to the Swiss Federal Institute of Technology in Lausanne. According to [27] and [28] the local port 23753/UDP is used to play a MMORPG, Heroes of Might and Magic. This can also be a hint of a gaming network services, like Microsoft's Direct Play for playing directX games over a network. As described in [28], the technology does open ports via UPnP.

We can observe, that some of the potential benign P2P traffic is originating by three different AS: AS3215, AS3269 and AS3352 belonging to the item-sets {**#35**}. As we can lookup in the ripe database [31], these AS belong to ripe itself, so this information does not bring any clarity on the cause of these flows. For further information of AS number assignment, try the link supplied in the registrar info of ripe.

Item-sets {**#36**} with the remote port 31415 belong to the XBSlink protocol for Xbox 360 and PS2/3 as described in the analysis with 10% support. They also have the flowtype set to 10, indicating failing connections.

The IP's 82.130.102.161 and 82.130.102.218 in the item-sets belong to fake e-mule servers. We assumed that the usage of port 4246 in the item-sets {**#37**} belongs to the fake e-mule server as they appeared in a item-set {**#38**} with one of the IP's. Now another port can be seen in the item-sets {**#39**} with lower support, port 4254 UDP. Some e-mule servers reserve a port range for sharing their content. Also the *flowtype* set to 10 for flows with local port 4254 indicates failing connections to the server {**#40**}. This could also indicate misconfiguration of server or clients due to outdated server lists.

The last important item-set {**#41**} contains the IP 134.21.2.59 with the local port 17543. The IP refers to the university of Fribourg [29]. According to the paper: „Identification and Analysis of Peer-to-Peer Traffic “ [30], the port 17543 is used by the application MP2P.

In the analysis with a support of 5%, no conclusion about the percentage of one way flows covered can be made. This is caused by splitting of item-sets into finer grained item-sets with a lesser percentage, resulting in not occurring item-sets because they have a support lesser than 5%.

5.10. Results Class Unreachable (Support 10%)

Tables 22 and 23 show the resulting item-sets of the class Unreachable with a support of 10%. The items *ASlocal:559* and *flowtype:2* are not shown, because they appear in every item-set. The *dPkts* and *durMs* attributes are only filtered in tables 22 and 23, when they are not needed to distinguish between item-sets.

File	Item-sets	Pct
flows_aug2008	prot:UDP tos:0	10.1%
	#42 dOctets:304 rPort:123 IIP:129.132.2.21 lPort:123 prot:UDP tos:0	10.3%
	#42, #43 dOctets:76 rPort:234 tos:0 IIP129.132.2.21 lPort:123 prot:UDP	11.3%
	#42 dOctets:380 rPort:123 tos:0 IIP:129.132.2.21 lPort:123 prot:UDP	39.9%
signs_aug2008	**#44** Onepkt UDP Unreach	24%

Table 22: Resulting item-sets of class Unreachable with 10% support in august 2008

File	Item-sets	Pct
flows_feb 2009	dPkts:3 prot:UDP tos:0	10.4%
	dPkts:1 prot:UDP tos:0 durMs:0	10.6%
	#42 dOctets:152 IIP:129.32.2.21 tos:0 lPort:123 prot:UDP	11%
	#42 dOctets:760 rPort:123 tos:0 IIP:129.132.2.21 lPort:123 prot:UDP	25.1%
signs_feb 2009	**#44** Onepkt UDP Unreach	10.9%

Table 23: Resulting item-sets of class Unreachable with 10% support in February 2009

5.10.1. Conclusions

Some of the resulting flow item-sets {**#42**} in table 22 and 23 use port 123 which is assigned to the Network Time Protocol (NTP). Remote hosts can send a request, which consists out of a UDP datagram without any payload. The response is a UDP datagram with the elapsed time in Milli seconds since the first of January 1900. As we can see a lot of clients are sending requests to the IP 129.132.2.21 which is the NTP server of the ETH (swisstime.ethz.ch). Flows trying to connect to the NTP server are assigned to this class, because the service was unreachable due to maintenance reasons or it did not respond due to malformed requests.

Considering the found item-sets, we can say that the rules for the class unreachable are well defined, because the NTP server is in a list containing all known hosts which distribute a service. For gaining the sign *Unreach*, flows are verified towards this list of known hosts.

One item-set {**#43**} shows the use of the port number 234/UDP. At the moment, nothing about this port is known. The IANA database only shows, that this port is reserved, but it is not specified for what, or for which application.

The not mentioned item-sets are not very descriptive. The items in the sets can occur on a variety of different one way flows. Thus it is not possible to gain information over the flows matching these item-sets.

Analyzing the sign item-set {**#44**} of both periods shows that they are the same. They only differ in their percentage of appearance. Flows matching these sign item-sets are mainly failing connection attempts, identified by the *Onepkt* sign and the Unreach sign, which indicates, that these host pairings had bidirectional connections before.

5.11. Class Unreachable (Support 5%)

The flowtype and *ASlocal* items are filtered. Furthermore the *ASremote* item is also filtered, because all flow item-sets (excluding the one formatted in italic) have the item *ASremote:5432* in the item-set. The one formatted in italic does not provide any *ASremote* attribute. Tables 24 and 25 show the resulting item-sets.

File	Item-sets	Pct
Flows aug2008	dOctets:228 dPkts:3 rPort:123 IIP:129.132.2.21 lPort:123 prot:UDP tos:0	7.0
	dOctets:152 dPkts:2 rPort:123 IIP:129.132.2.21 lPort:123 prot:UDP tos:0	6.7
	dOctets:304 dPkts:4 rPort:123 IIP:129.132.2.21 lPort:123 prot:UDP tos:0	10.3
	dOctets:76 rPort:123 tos:0 dPkts:1 durMs:0 IIP:129.132.2.21 lPort:123 prot:UDP	11.3
	dOctets:380 dPkts:5 rPort:123 tos:0 IIP:129.132.2.21 lPort:123 prot:UDP	39.9
Signs aug2008	**#45** TRWnom UDP Unreach	5.3
	Onepkt UDP Unreach	24.0

Table 24: Resulting item-sets of class Unreachable with 5% support in august 2008

File	Item-sets	Pct
Flows feb2009	dOctets:456 dPkts:6 rPort:123 tos:0 IIP:129.132.2.21 lPort:123 prot:UDP	5.2
	dOctets:380 dPkts:5 rPort:123 IIP:129.132.2.21 lPort:123 prot:UDP tos:0	6.3
	dOctets:608 rPort:123 tos:0 dPkts:8 IIP:129.132.2.21 lPort:123 prot:UDP	6.6
	dOctets:304 dPkts:4 rPort:123 IIP:129.132.2.21 lPort:123 prot:UDP tos:0	7.8
	dOctets:76 IIP:129.132.2.21 tos:0 dPkts:1 durMs:0 lPort:123 prot:UDP	8.0
	dOctets:228 dPkts:3 rPort:123 IIP:129.132.2.21 lPort:123 prot:UDP tos:0	8.6
	dOctets:152 dPkts:2 rPort:123 tos:0 IIP:129.132.2.21 lPort:123 prot:UDP	8.1
	dOctets:760 dPkts:10 rPort:123 tos:0 IIP:129.132.2.21 lPort:123 prot:UDP	25.1
Signs feb2009	Onepkt UDP Unreach	10.90%

Table 25: Resulting item-sets of class Unreachable with 5% support in February 2009

5.11.1. Conclusions

Analyzing this class with a smaller support does not reveal any information as shown in tables 24 and 25. This is because all flow item-sets are directed to the timeserver swisstime.ethz.ch, what has already been discovered in the analysis of this class with a support of 10%. This fact can lead to the conclusion, that the rules for this class are well defined. But with these results, the peak of the class in the analyzed period was not inspected.

The sign item-set {**#45**} listed in table 24 could indicate a malicious scanning activity, because it has the item *TRWnom* set. The other sign item-sets are the same as in the analysis with a support larger than 10%.

Inspecting the percentage of covered flows leads to the conclusion, that all flows having the item IIP:129.132.2.21 set matching to this class, because the host with this IP is covered by the file containing all known hosts with their according service.

6. Analyzing reference intervals

6.1. Purpose

Analyzing a reference interval is important to understand, which flow item-sets can also be discovered in a period without peak. Then we can say which flow and sign item-sets are new in the period with peak. This also helps in identifying, if the chosen interval is accurate or if we have to make an analysis over a whole period to determine the cause of a peak.

6.2. Analysis setup

For this analysis, some corrections in the `sign_eval2_for_FIM tool` is made. Therefore the local AS is not shown anymore, because it is always the same AS. Also some improvements in writing the analyzed signs and flows are made, so the output files can be read directly by the FIM tool `sam`. The analysis is done with a support of 5%, so the observation of flow and sign item-sets are finer grained.

To have reference item-sets, a interval of 3 continuous hours is chosen out of 4 previous periods without a significant peak. Table 26 lists those interval and the according time periods per class.

Class	Period	Interval
Other Malicious	August 2006	30.07.2006 01.00 to 04.00
	February 2007	28.01.2007 19.00 to 23.00
	August 2007	31.07.2007 05.00 to 08.00
	February 2008	03.02.2008 04.00 to 07.00
Backscatter	February 2005	31.01.2005 02.00 to 05.00
	August 2005	29.07.2005 06.00 to 09.00
	February 2006	01.02.2006 09.50 to 12.50
	August 2006	30.07.2006 01.00 to 04.00
Unreachable	August 2006	29.07.2006 22.00 to 30.07.2006 01.00
	February 2007	28.01.2007 16.00 to 19.00
	August 2007	31.07.2007 02.00 to 05.00
	February 2008	03.02.2008 04.00 to 07.00
Benign P2P	August 2005	29.07.2005 06.00 to 09.00
	February 2006	01.02.2006 09.50 to 12.50
	August 2006	30.07.2006 01.00 to 04.00
	February 2007	28.01.2007 19.00 to 22.00

Table 26: Listing of reference period and chosen interval

The following results do not document the *flowtype:2* and *ASlocal:559* items for the item-sets, because they are always present. The flowtype is only listed, if it does not have the value 2. For better displaying purposes, some flow attributes are not listed, depending on the space a flow item-set needs. If any information is truncated, it will be mentioned at the top of the result section.

The location of the resulting output files are listed in the appendix.

6.3. Results Class Other Malicious

6.3.1. Results August 2006

The results in table 27 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

Sort	Item-sets	Pct
Flows	**#46** rPort:123 lPort:123 prot:UDP	5.4 %
	#47 ASremote:8404 lPort:53 prot:UDP	6.2 %
	#46 lIP:130.60.7.44 lPort:123 tos:0 prot:UDP	5.2 %
	#48 rIP:218.85.139.179 rPort:9001 ASremote:4134 tos:128 prot:TCP	7.4 %
	#46 lIP:130.60.7.43 lPort:123 tos:0 prot:UDP	6.3 %
	#49 lPort:37 lIP:129.132.2.21 prot:TCP	7.3 %
	#46 lIP:130.60.7.52 lPort:123 tos:0 prot:UDP	7.5%
Signs	**#51** TRWnom GreyIP TCP	7.8 %
	#51 TRWnom UDP Onepkt	11.3 %
	GreyIP UDP	5.5 %
	#50 TCP Onepkt	11.5 %

Table 27: Listing of observed flow and sign item-sets from interval in period August 2006

The item-set {**#46**} in table 27 describes traffic directed to a NTP server, because the item-sets use the local port 123/UDP. The IP addresses shown in three of the four item-sets {**#46**} are NTP servers from ch.pool.ntp.org and are in the IP space of the University of Zurich.

Item-set {**#47**} describes one-way flows to the port 53/UDP, which is assigned to the DNS service.

Looking up the ripe database allocates the IP in item-set {**#48**} to China Telecom. The remote port 9001 over UDP is used by various services and applications as described in [32]. IANA assigned the port to the ETL Service Manager [33] and also the tor applications is using this port, when acting as a proxy [34].

The last item-set to describe is {**#49**}. This item-set marks failing connection attempts to the server swisstime.ethz.ch, trying to use the obsolete Time Protocol. The server only serves the newer NTP protocol.

Failing connection attempts could be the cause of all flows having the sign item-set {**#50**}. The cause could be a client, trying to connect to a server, which does not respond, or the packet is dropped by a firewall or due to congestion of the network.

It can be observed, that two item-sets {**#51**} have the *TRWnom* sign set. This could indicate slow or stealth scanning or, in the case of TCP, Half SYN scans.

6.3.2. Results February 2007

The results in table 29 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

The results in table 29 reduce the following item-sets to one single item-set {**#55**}. This is done, because the item-sets differ in the amount of packets and therefore are not overlapping. The original item-sets are shown in table 28.

Original item-sets
dPkts:6 lPort:4662 tos:0 prot:TCP
lPort:4662 dOctets:144 tos:0 dPkts:3 prot:TCP

Table 28: Orignal flow item-sets of period February 2007

Sort	Item-sets	Pct
Flows	**#52** rIP:62.2.24.162 ASremote:8404 lPort:53 prot:UDP tos:0	5.0 %
	#52 rIP:62.2.17.61 ASremote:8404 lPort:53 prot:UDP tos:0	5.8 %
	#52 rIP:62.2.17.60 ASremote:8404 lPort:53 prot:UDP tos:0	6.3 %
	#53 lIP:82.130.70.8 ASremote:8404 lPort:53 prot:UDP tos:0	5.9 %
	#53 lIP:82.130.116.11 ASremote:8404 lPort:53 prot:UDP tos:0	5.9 %
	#53 lIP:82.130.70.6 ASremote:8404 lPort:53 prot:UDP tos:0	5.9 %
	#54 ASremote:8404 lPort:53 prot:UDP tos:0	7.4 %
	#55 lPort:4662 tos:0 prot:TCP	13 %
Signs	**#56** P2P TCP GreyIP	27.5 %
	#57 TRWnom Onepkt UDP	27.6 %
	UDP GreyIP	6.7 %

Table 29: Flow and sign item-sets of the interval out of period February 2007

Cablecom is a swiss regional ISP and their IP space assigned is 62.2.0.0/16. The contacted port of the item-sets {**#52**} having originating IP's assigned to Cablecom is using the protocol UDP, which indicates DNS lookup traffic.

The IP space 82.130.64.0/18 shown in item-set {**#53**} is assigned to ETH Zurich. Looking at the used port indicates DNS lookup traffic. Another item-set {**#54**} having the local port set to 53 using UDP can be observed. With the given facts, this flow item-set describes DNS lookup traffic.

Observing the local port set to 4662 and communicating over TCP in the item-set {**#55**} leads to the conclusion, that this is P2P traffic directed to a e-mule server or client. E-mule uses TCP to exchange parts of a file.

Item-set {**#56**} can result from outdated server list, indicated by the *GreyIP* sign. This could be a client having an outdated server list and trying to connect to a IP address, which no longer exists.

Having the *TRWnom* sign set in item-set {**#57**}, could indicate stealth or slow scanning activities. But this is hard to tell, whether the activities are benign or malicious, without any other informations.

6.3.3. Results August 2007

The results in table 30 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

Sort	Item-sets	Pct
Flows	**#58** lPort:4662 prot:TCP	5.6 %
	#59 lIP:129.132.97.15 tos:0 lPort:123 prot:UDP	8.6 %
	#60 ASremote:4134 lPort:37 tos:128 lIP:129.132.2.21 prot:UDP	36.4 %
Signs	TRWnom	5.2 %
	#61 P2P TCP GreyIP	9.7 %
	Onepkt UDP	63.4 %

Table 30: Listing of flow and sign item-sets found in the interval of period August 2007

Table 30 shows the item-sets of the interval analyzed in period August 2007. Item-set {**#58**} describes connections to a e-mule client or server by using the local port 4662. Flows directed to a NTP server are described by item-set {**#59**}. The last item-set {**#60**} describes the failing connection attempts to the time server swisstime.ethz.ch, which only provides NTP time services and not the obsolete Time Service protocol.

Potentially malicious P2P traffic is assigned to this interval as shown in item-set {**#61**}. The flows are not classified in the class benign P2P, because they have the *GreyIP* sign set.

The other two sign item-sets are not very informative. Only the percentage of the item-set {**#62**} indicates a lot of flows only having one packet, which can have a lot of benign and malicious causes.

6.3.4. Results February 2008

The results in table 31 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets. Furthermore the item-set {**#63**} does not list the *ASremote:3143* attribute so it can be fitted to one single line.

File	Item-sets	Pct
Flows	**#63** dOctets:32 tos:128 lPort:37 lIP:129.132.2.21 dPkts:1 durMs:0 prot:UDP	68.2
Signs	**#64** TCP GreyIP **#65** Onepkt UDP	10.7% 83.6%

Table 31: Results from reference interval of class Other Malicious in

Table 31 shows, that the time server from ETH {**#63**} is the largest item-set, which corresponds to the findings in our previous analysis. The other flow item-sets are not very informative, we can only observe, that the two main transport protocols TCP and UDP are present.

Taking a look at the sign item-sets shows that the item-set {**#64**} has a larger support than in the inspected interval of the peak periods. It's amount is almost the double.

Another observation is that the item-set {**#65**} has slight smaller occurrence but is also massively present.

This inspected interval shows, that some item-sets are missing:

- *GreyIP, UDP*
- *TRWnom, UDP*
- *TRWnom, Onepkt*

Not observing these item-sets in the results of the reference period above, is an indicator of what traffic could have caused the peak, because two of the missing item-sets have the sign *TRWnom* set, which is an indicator for slow scanning activities. Of course, this conclusion can only be approved, if we inspect other periods for the occurrence of item-sets with the *TRWnom* sign set.

6.4. Results Class Backscatter

6.4.1. Results February 2005

The results in table 33 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

Because the truncation of the mentioned flow item-set attributes, some item-sets look identical. These item-sets are reduced and represented as a single item-set. The reduction is possible, because the item-sets differ in the attribute *dOctets* or the *dPkts* value and therefore are not overlapping. The percentage of appearance is added up. Table 32 shows the original and the combined item-sets.

Original Item-sets	Combined Item-set
dOctets:180 dPkts:3 lPort:0 tos:0 rPort:0 prot:ICMP dOctets:46 lPort:0 dPkts:1 durMs:0 tos:0 rPort:0 prot:ICMP dOctets:56 dPkts:1 lPort:0 durMs:0 tos:0 rPort:0 prot:ICMP dOctets:28 durMs:0 lPort:0 dPkts:1 tos:0 rPort:0 prot:ICMP	lPort:0 tos:0 rPort:0 prot:ICMP
dOctets:79 rIP:61.178.183.216 ASremote:4134 dPkts:2 lPort:0 tos:0 rPort:0 prot:ICMP	rIP:61.178.183.216 ASremote:4134 lPort:0 tos:0 rPort:0 prot:ICMP
dOctets:33 rIP:61.178.183.216 ASremote:4134 dPkts:1 durMs:0 lPort:0 tos:0 rPort:0 prot:ICMP	

Table 32: Combined item-sets in February 2005

Sort	Item-sets	Pct
Flows	**#66** rIP:61.178.183.216 ASremote:4134 lPort:0 tos:0 rPort:0 prot:ICMP **#67** rIP:62.139.133.18 dOctets:74 ASremote:20858 lPort:0 tos:0 rPort:0 prot:ICMP lPort:0 tos:0 rPort:0 prot:ICMP **#68** rIP:67.18.109.122 ASremote:21844 lPort:0 tos:0 rPort:0 prot:ICMP	15.3 % 5.7 % 36.5 % 9.7 %
Signs	**#69** GreyIP Onepkt ICMP	25.4 %

Table 33: Flow and sign item-sets discovered in interval of period Febrary 2005

Table 33 lists the discovered item-sets. All item-sets have the protocol ICMP and their *rPort* value set to 0, which indicates, that the router does not log ICMP type and code. The owner of the listed IP's are somehow suspicious, because the IP space 61.178.183.0/24 in item-set {**#66**} is assigned to one person according to the APNIC [35]. IP {**#67**} belongs to EGYNET-DSL an Egyptian ISP. According to ARIN the third IP in item-set {**#68**} belongs to an ISP located in the area of Houston Texas [36].

The sign item-set{**#69**} displayed in table 33 is very general. It can not be determined, to which rule it belongs.

6.4.2. Results August 2005

The results in table 35 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

Because the truncation of the mentioned flow item-set attributes, some item-sets look identical. These item-sets are reduced and represented as a single item-set. The reduction is possible, because the item-sets differ in the attribute *dOctets* or the *dPkts* value and therefore are not overlapping. The percentage of appearance is added up. Table 34 shows the original and the combined item-sets.

Original item-sets	Combined item-sets
dPkts:3 tos:0 lPort:0 prot:ICMP dPkts:2 tos:0 lPort:0 prot:ICMP	tos:0 lPort:0 prot:ICMP
dOctets:46 rPort:0 dPkts:1 tos:0 durMs:0 lPort:0 prot:ICMP dOctets:56 rPort:0 dPkts:1 tos:0 durMs:0 lPort:0 prot:ICMP	rPort:0 tos:0 lPort:0 prot:ICMP
dOctets:92 rPort:8 tos:0 lPort:0 prot:ICMP dOctets:46 rPort:8 tos:0 dPkts:1 durMs:0 lPort:0 prot:ICMP	rPort:8 tos:0 lPort:0 prot:ICMP

Table 34: Listing of original and reduced item-sets in August 2005

Sort	Item-sets	Pct
Flows	lPort:0 prot:ICMP	5.1 %
	tos:0 lPort:0 prot:ICMP	10.4 %
	#72 rPort:771 lPort:0 prot:ICMP	5.6 %
	ASremote:4134 tos:128 lPort:0 prot:ICMP	5.6 %
	#71 rPort:11 lPort:0 prot:ICMP	5.9 %
	#71 rPort:11 tos:0 lPort:0 prot:ICMP	5.3 %
	#70 rIP:152.98.224.108 rPort:0 lPort:0 Asremote:7575 tos:0 prot:ICMP	11.1 %
	#71 rPort:8 tos:0 lPort:0 prot:ICMP	13.2 %
	#70 rPort:0 tos:0 lPort:0 prot:ICMP	13.5 %
	tos:128 lPort:0 prot:ICMP	6.1 %
	#71 tos:128 rPort:8 lPort:0 prot:ICMP	9.7 %
	#71 rPort:8 lPort:0 prot:ICMP	8.5 %
Signs	GreyIP Onepkt ICMP	47 %

Table 35: Flow and sign item-sets gathered in period August 2005

Two item-sets {**#70**} with the *rPort* value set to 0 can be observed in table 35. One of them has an IP address in it belonging to University of Queensland. The table also lists item-sets {**#71**} having their *rPort* value set to 11 or 8. These ICMP messages are not defined.

Item-set {**#72**} lists the *rPort* value 771 which is a port unreachable of type 0x03 and code 0x03.

The sign item-set shown in table 35 is not very informative. It only describes a lot of flows which are targeting an unused local IP address.

6.4.3. Results February 2006

The results in table 37 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

Because the truncation of the mentioned flow item-set attributes, some item-sets look identical. These item-sets are reduced and represented as a single item-set. The reduction is possible, because the item-sets differ in the attribute *dOctets* or the *dPkts* value and therefore are not overlapping. The percentage of appearance is added up. Table 36 shows the original and the combined item-sets.

Original item-sets	Reduced item-sets
dOctets:92 rPort:8 durMs:0 tos:0 lPort:0 prot:ICMP dOctets:46 rPort:8 durMs:0 tos:0 dPkts:1 lPort:0 prot:ICMP	rPort:8 tos:0 lPort:0 prot:ICMP
dOctets:92 dPkts:1 durMs:0 tos:0 lPort:0 prot:ICMP dOctets:112 dPkts:2 tos:0 lPort:0 prot:ICMP	tos:0 lPort:0 prot:ICMP
tos:128 dPkts:2 lPort:0 prot:ICMP tos:128 dPkts:1 lPort:0 durMs:0 prot:ICMP	lPort:0 prot:ICMP

Table 36: Original and reduced flow item-sets in period February 2006

Sort	Item-sets	Pct
Flows	tos:192 lPort:0 prot:ICMP	5.4 %
	#74 rPort:0	5.0 %
	#74 rPort:0 tos:0 lPort:0 prot:ICMP	7.1 %
	#72 rPort:8 lPort:0 prot:ICMP	7.3 %
	#72 tos:128 rPort:8 lPort:0 prot:ICMP	7.8 %
	#72 rPort:8 tos:0 lPort:0 prot:ICMP	12.7 %
	#72 dPkts: 2 rPort:8 tos:0 lPort:0 prot:ICMP	6.4 %
	tos:0 lPort:0 prot:ICMP	11.7 %
	lPort:0 prot:ICMP	15.3 %
	#72 rPort:11 lPort:0 prot:ICMP	5.2 %
	#72 rPort:11 tos:0 lPort:0 prot:ICMP	11.4 %
	#73 rPort:771 tos:0 lPort:0 prot:ICMP	6.8 %
	#73 tos:128 rPort:771 lPort:0 prot:ICMP	6.8 %
	#73 rPort:771 lPort:0 prot:ICMP	5.7 %
Signs	**#75** Large ICMP	9.8 %
	GreyIP Onepkt ICMP	20.9 %

Table 37: Flow and sign item-sets in period February 2006

Table 37 shows a lot of item-sets {**#72**} with non existent ICMP type and code information in their *rPort* values. Item-sets {**#73**} representing an ICMP port unreachable message are also present. Some item-sets {**#74**} have their *rPort* value set to 0 which indicates, that the router does not log the ICMP type and code informations.

The sign item-set {**#75**} indicates ICMP flows with an amount of packets larger 9 or at least 10240 bytes of size. ICMP messages are not very long, so the Large sign could indicate a host under a ICMP flood. ICMP messages can also be used to tunnel data via the ICMP payload, which is another explanation for the occurrence of the Large sign.

The other item-set in table 37 is a returning one, which can be observed in all reference intervals of the class Backscatter.

6.4.4. Results August 2006

The results in table 39 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

Because the truncation of the mentioned flow item-set attributes, some item-sets look identical. These item-sets are reduced and represented as a single item-set. The reduction is possible, because the item-sets differ in the attribute *dOctets* or the *dPkts* value and therefore are not overlapping. The percentage of appearance is added up. Table 38 shows the original and the combined item-sets.

Original item-sets	Reduced item-sets
dOctets:92 rPort:8 dPkts:1 tos:0 durMs:0 lPort:0 prot:ICMP dPkts:2 rPort:8 tos:0 lPort:0 prot:ICMP	rPort:8 tos:0 lPort:0 prot:ICMP
dOctets:112 rPort:11 dPkts:2 lPort:0 prot:ICMP rPort:11 dOctets:56 dPkts:1 durMs:0 lPort:0 prot:ICMP	rPort:11 lPort:0 prot:ICMP
dOctets:112 rPort:771 tos:0 lPort:0 prot:ICMP dOctets:56 rPort:771 tos:0 dPkts:1 durMs:0 lPort:0 prot:ICMP	rPort:771 tos:0 lPort:0 prot:ICMP

Table 38: Original and reduced flow item-sets in period August 2006

One of the item-sets {**#77**} in table 39 displays the *dPkts:2* attribute. The attribute not removed to distinguish this item-set from the merged item-sets having the *rPort* value set to 771. The distinction is made, because the item-sets might overlap.

Sort	Item-sets	Pct
flows_aug2006	**#78** IIP:192.33.90.198 rPort:11 lPort:0 prot:ICMP	5.1 %
	ASremote:4837 lPort:0 prot:ICMP	6.6 %
	tos:0 lPort:0 prot:ICMP	7.9 %
	tos:128 lPort:0 prot:ICMP	6.5 %
	ASremote:4134 tos:128 lPort:0 prot:ICMP	5.9 %
	#76 rPort:8 tos:0 lPort:0 prot:ICMP	15.9
	#76 tos:128 rPort:8 lPort:0 prot:ICMP	6.7 %
	#76 rPort:11 dPkts:2 lPort:0 prot:ICMP	15.6 %
	#76 rPort:11 tos:0 lPort:0 prot:ICM	10.9 %
	#76 tos:128 rPort:11 lPort:0 prot:ICMP	6.1 %
	#77 rPort:771 tos:0 lPort:0 prot:ICMP	12 %
	#77 dPkts:2 rPort:771 tos:0 lPort:0 prot:ICMP	6.6 %
	#77 tos:128 rPort:771 lPort:0 prot:ICMP	6.1 %
signs_aug2006	**#79** Large ICMP	6.6%
	#79 GreyIP Onepkt ICMP	19.1%

Table 39: Results from reference interval of class Other Malicious in August 2006

Table 39 shows that some item-sets shown have the same ICMP message types as observed in the FIM analysis of the class backscatter with 10%. The item-sets {**#76**} having their *rPort* set to 8 or 11 are not defined as ICMP message types having the type both set to 0 and the code to 8 respectively to 11. A observation of a valid ICMP type and code is when the *rPort* value is set to 771 {**#77**}, indicating a port unreachable ICMP message.

The IP address 192.33.90.198 {**#78**} belongs to the Distributed Computing Group from ETH [37], which are doing research on distributing service and P2P behavior.

Both of the sign item-sets {**#79**} do also appear in the analysis of the interval form a peak period. The only anomaly found here is the missing item-set (*TRWnom*, *GreyIP*, *ICMP*). The *TRWnom* sign of the item-set found in the peak period could indicate a slow scanning activities. But this can only be approved, if all intervals are inspected, because the item-set could only be missing in the reference interval but could be observed in another non peak period.

6.5. Results Class Benign P2P

6.5.1. Results August 2005

The results in table 41 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

Because the truncation of the mentioned flow item-set attributes, some item-sets look identical. These item-sets are reduced and represented as a single item-set. The reduction is possible, because the item-sets differ in the attribute *dOctets* or the *dPkts* value and therefore are not overlapping. The percentage of appearance is added up. Table 40 shows the original and the combined item-sets.

Original item-sets	Reduced item-set
dOctets:38 IIP:129.132.57.4 durMs:0 tos:0 dPkts:1 prot:UDP IIP:129.132.57.4 dPkts:1 durMs:0 prot:UDP tos:0	IIP:129.132.57.4 prot:UDP tos:0

Table 40: Original and reduced item-sets in period August 2005

Sort	Item-sets	Pct
Flows	**#82** IIP:129.132.73.145	5.2 %
	#83 IIP:195.176.0.50	5.2 %
	#81 rPort:4672 prot:UDP	5.8 %
	#81 lPort:4672 dPkts:1 prot:UDP	5.3 %
	#81 lPort:4672 prot:UDP	5.5 %
	#82 IIP:129.132.57.4 tos:0 prot:UDP	12.2 %
	#80 rPort:6881 tos:0 prot:UDP	5.4 %
	#80 rPort:6881 lPort:6881 prot:UDP	5.1 %
	#80 rPort:6881 lPort:6881 prot:UDP tos:0	5.6 %
	#80 lPort:6881 tos:0 prot:UDP	7.1 %
	flowtype:10 prot:TCP	5.9 %
	flowtype:10 prot:UDP tos:0	5.2 %
	#80 lPort:6881 prot:TCP tos:0	5.2 %
Signs	**#84** PotOk TCP P2P	5.9 %
	#84, #85 PotOk Onepkt UDP P2P	5.5 %

Table 41: Flow and sign item-sets in period August 2005

Some of the item-sets {**#80**} displayed in table 41 use the remote port 6881 on TCP and UDP. This port is used by bit torrent applications. Some of the flows described within these item-sets also use the local port 6881 over UDP and TCP.

Item-sets {**#81**} with the remote or local port set to 4672/UDP are e-mule clients. This port is used by e-mule for client to client communications.

The IP addresses in the item-sets {**#82**} are assigned to the ETH Zurich. The IP address in item-set {**#83**} belongs to the uplink network [38] which connections student houses to the internet.

Both of the occuring sign item-sets {**#84**} mark potentially bening P2P traffic over both common layer 4 protocols. The sign item-set {**#85**} describes flows, only consisting out of one packet. This could indicate traffic to other P2P clients which exchanges peer informations.

6.5.2. Results February 2006

The results in table 43 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

Because the truncation of the mentioned flow item-set attributes, some item-sets look identical. These item-sets are reduced and represented as a single item-set. The reduction is possible, because the item-sets differ in the attribute value *dPkts* and therefore are not overlapping. The percentage of appearance is added up. Table 42 shows the original and the combined item-sets.

Original item-sets	Reduced item-sets
dPkts:3 lPort:6881 dPkts:2 lPort:6881 flowtype:2	lPort:6881

Table 42: Listing of the original and reduced flow item-sets in period February 2006

Sort	Item-sets	Pct
Flows	flowtype:10 tos:0	6.2 %
	#86 lPort:6881	10.6 %
	#86 rPort:6881 lPort:6881 prot:UDP	6.0 %
	#86 rPort:6881 tos:0 prot:UDP	8.3 %
	#86 rPort:6881 lPort:6881 tos:0 prot:UDP	6.5 %
	#86 tos:128 lPort:6881	5.2 %
	#86 lPort:6881 tos:0 prot:UDP	20.4 %
	#86 prot:TCP lPort:6881 tos:0	6.2 %
Signs	**#87** PotOk P2P	7.5 %
	#87, #88 TCP P2P	26.8 %
	#87 Onepkt UDP P2P	45.9%

Table 43: Flow and sign item-sets found by inspecting interval of period February 2006

Almost all of the item-sets {**#86**} displayed in table 43 use the local or remote port 6881. Some over TCP, some over UDP. These item-sets mark flows belonging to the P2P application bit torrent.

It can be observed, that the sign item-sets {**#87**} shown in table 43 are splited up from the item-sets shown in table 41. We mark that the item-set {**#88**} containing the *TCP* sign does not mark potentially benign traffic anymore.

6.5.3. Results August 2006

The results in table 44 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

Sort	Item-sets	Pct
Flows	**#89** lPort:6881 prot:UDP	5.8 %
	#89 lPort:6881 tos:0	6 %
	#90 flowtype:10	5.2 %
	#90 flowtype:10 prot:UDP tos:0	10.5 %
Signs	**#91** Large UDP P2P	5.1 %
	#91 TRWnom PotOk UDP P2P	8.8 %
	#91 TCP P2P	29.6 %
	#91 Onepkt UDP P2P	33.2 %

Table 44: Discovered flow and sign item-sets in period August 2006

Table 44 shows the gained item-sets in the interval of period August 2006. The item-sets {**#89**} using the port number 6881 refer to a bit torrent application. Item-sets {**#90**} containing the *flowtype:10* are not very interesting, because they mark only failing connections between two hosts having benign connections otherwise.

The sign item-sets {**#91**} do not provide much information. They discribe benign P2P traffic.

6.5.4. Results February 2007

The results in table 45 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets.

File	Item-sets	Pct
flows_feb 2007	dOctets:182 dPkts:2 prot:UDP	5.5%
	#92, #95 lPort:27273 dPkts:2 tos:0 flowtype:10 prot:UDP	6.3%
	dOctets:288 dPkts:6 prot:TCP tos:0	5.6%
	#93 lPort:6881 tos:0	6.0%
	#94 lPort:4662 tos:0 prot:TCP	8.2%
	dOctets:91 tos:0 dPkts:1 durMs:0 prot:UDP	6.8%
	dOctets:144 dPkts:3 prot:TCP tos:0	7.3%
	tos:128 dPkts:1 prot:UDP durMs:0	8.3%
	tos:128 prot:TCP	6.3%
	dPkts:2 prot:UDP tos:0	14.7%
signs_feb 2007	**#96** TRWnom PotOk UDP P2P	5.0%
	#98 TCP P2P	35.7%
	#97 Onepkt UDP P2P	28.3%

Table 45: Results from reference interval of class Benign P2P

In the results of the reference interval, listed in table 45, we can observe three flow item-sets describing benign P2P traffic. The first has the local port 27273 {**#92**}. The local port 6881 {**#93**} is the default port used by bit Torrent clients, where as the local port 4662 {**#94**} is used by e-mule clients, described in [30].

One item-set {**#95**} has the flowtype 10, indicating failing connections between a host pairing which have otherwise established connections.

Another observation is the absence of IP's corresponding to fake e-mule server. In these terms, we also do not have any item-set having the items *lPort* or *rPort* set to 4246 or 4254. This hardens the implication, that these ports are associated to the fake e-mule server. Not observing these item-sets does not mean, they are not present. They probably only have a occurrence of less than 5%.

Taking a look at the sign item-sets shows that the item-sets observed in the reference period are slightly different. The item-set {**#96**} does not have the item *Onepkt* in it. This is no anomaly if we see, that a item-set {**#97**} with a much higher percentage exists, having this sign set, with two others of the item-set {**#96**}.

Interesting is the fact, that the item-set {**#98**} in the reference interval has a much higher occurrence than in the peak interval analyzed.

6.6. Results Class Unreachable

6.6.1. Results August 2006

The results in table 46 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets. The flow item-sets printed in italic font do not display the *ASremote:21494* attribute.

Sort	Item-sets	Pct
Flows	**#99, #106** IIP:195.176.255.135 tos:0 lPort:80 prot:TCP	5.4 %
	#100 rIP:81.221.252.10 ASremote:21494 lPort:53 prot:UDP	8.2 %
	#102 IIP:129.132.2.21 tos:0 lPort:123 prot:UDP	5.0 %
	#100, #105 IIP:195.176.20.204 prot:UDP lPort:53	5.6 %
	#100, #105 IIP:195.176.20.204 lPort:53 prot:UDP tos:0	5.9 %
	#103 lPort:25 flowtype:10 prot:TCP	5.3 %
	#103 lPort:25 tos:0 prot:TCP	6.9 %
	#99 ASremote:8404 lPort:80 tos:0 prot:TCP	9.4 %
	#100, #101 *rIP:81.221.250.10 IIP:130.82.128.1 lPort:53 prot:UDP*	5.4 %
	#100 #101 *rIP:81.221.250.10 IIP:130.82.128.2 lPort:53 prot:UDP*	5.4 %
	flowtype:10 prot:TCP tos:0	5.4 %
	#100 lPort:53 prot:UDP,tos:0	6.8 %
	#99, #104 rIP:85.3.195.113 ASremote:3303 IIP:160.85.36.247 lPort:80 prot:TCP	17.4 %
Signs	**#107** Large TCP Unreach	10.2 %
	#110 P2P TCP Unreach	5.8 %
	#110 P2P UDP Unreach	5.5 %
	#108 TRWnom PotOk UDP Unreach	5.9 %
	#108 TRWnom UDP Onepkt Unreach	11.9 %
	PotOk UDP Onepkt Unreach	14.5 %
	#109 TCP Onepkt Unreach	20.5 %

Table 46: Flow and sign item-sets discovered in period August 2006

Some one way flows directed to a web server can be observed in the flow item-sets {**#99**} shown in table 46. Another protocol is present in the results, using remote and/or local port 53 on UDP {**#100**} which is assigned to the DNS service. DNS servers use the UDP protocol to send and receive messages to clients requesting a DNS lookup. A lot of traffic from a specific IP, as listed in the item-sets {**#101**} could be interpreted as a DoS attack or a service having a wrong IP address configured for looking up domain names. By inspecting the IP addresses more preciser, we see that 81.221.250.10 belongs to green.ch, a swiss customer ISP. The IP addresses 130.82.128.1 and 130.82.128.2 belong to the university of St. Gallen. Having this evidence leads to the conclusion of contacting invalid DNS servers due to configuration mistakes or topology changes.

One item-set {**#102**} refers to the time server from the ETH Zurich (swisstime.ethz.ch). Two other item-sets {**#103**} are found, which have the item local port with the value 25 over the TCP transport protocol. Port 25 is assigned to the SMTP protocol, for transferring e-mails to the server.

The flow item-set {**#104**} describes a machine in the IP space of bluewin (85.3.195.113) trying to contact a web server located in the IP space of ZHAW Winterthur (160.85.26.247).

Looking up the IP address 195.176.20.204 {**#105**} in a ripe.net query, reveals that his IP address is located at the IBM Research Center in Rüschlikon. Using port 53 over UDP leads to the conclusion, that DNS lookups are performed by this machine.

There is another item-set {**#106**} with the IP 195.176.255.135, which is allocated in the IP range of SWITCH, communicating over TCP port 80, indicating web server traffic.

The sign item-sets shown in table 46 all mark traffic directed to a known server, which is failing. This is indicated by the Unreach sign. Item-set {**#107**} marks failing TCP traffic, which at least consists out of 10 packets or has a minimum size of 10240 bytes.

The *TRWnom* sign contained in these item-sets {**#108**} marks some delayed failing connection attempts. The last item-set {**#109**} discussed shows TCP connection attempts consisting out of a single packet. This indicates, that the first SYN packet was sent, but the opponent hasn't sent back a SYN ACK packet, due to network congestion or a firewall dropping the traffic due to its rules.

The sign item-sets {**#110**} indicating a P2P application trying to connect to a machine, which is no longer available. This behavior is observed, when some clients have outdated server lists.

6.6.2. Results February 2007

The results in table 47 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets. The flow item-sets printed in italic font do not display the *Asremote:21494* attribute.

Sort	Item-sets	Pct
Flows	**#99, #106** lIP:195.176.255.135 tos:0 lPort:80 prot:TCP	5.4 %
	#100 rIP:81.221.252.10 ASremote:21494 lPort:53 prot:UDP	8.2 %
	#102 lIP:129.132.2.21 tos:0 lPort:123 prot:UDP	5.0 %
	#100, #105 lIP:195.176.20.204 prot:UDP lPort:53	5.6 %
	#100, #105 lIP:195.176.20.204 lPort:53 prot:UDP tos:0	5.9 %
	#103 lPort:25 flowtype:10 prot:TCP	5.3 %
	#103 lPort:25 tos:0 prot:TCP	6.9 %
	#99 ASremote:8404 lPort:80 tos:0 prot:TCP	9.4 %
	#100, #101 *rIP:81.221.250.10 lIP:130.82.128.1 lPort:53 prot:UDP*	5.4 %
	#100 #101 *rIP:81.221.250.10 lIP:130.82.128.2 lPort:53 prot:UDP*	5.4 %
	flowtype:10 prot:TCP tos:0	5.4 %
	#100 lPort:53 prot:UDP,tos:0	6.8 %
	#99, #104 rIP:85.3.195.113 ASremote:3303 lIP:160.85.36.247 lPort:80 prot:TCP	17.4 %
Signs	**#111** PotOk UDP Onepkt Unreach	12.5 %
	#112 TCP Onepkt Unreach	14.9 %

Table 47: Flow and sign item-sets discovered in period February 2007

The flow item-sets displayed in table 47 are the same as in table 47. They also have exactly the same percentage in occurrence. Please refer to the explanations of table 47.

Item-set {**#111**} in listed in table 48 shows potentially benign UDP traffic, which has only one packet assigned. Applications based on UDP often send status information to inform a peer that its still alive.

Flows matching to item-set {**#112**} could be lost due to network congestion or be dropped from a firewall and therefore only having one packet.

6.6.3. Results August 2007

The results in table 48 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets. The flow item-sets printed in italic font do not display the *Asremote:21494* attribute.

Sort	Item-sets	Pct
Flows	**#99, #106** lIP:195.176.255.135 tos:0 lPort:80 prot:TCP	5.4 %
	#100 rIP:81.221.252.10 ASremote:21494 lPort:53 prot:UDP	8.2 %
	#102 lIP:129.132.2.21 tos:0 lPort:123 prot:UDP	5.0 %
	#100, #105 lIP:195.176.20.204 prot:UDP lPort:53	5.6 %
	#100, #105 lIP:195.176.20.204 lPort:53 prot:UDP tos:0	5.9 %
	#103 lPort:25 flowtype:10 prot:TCP	5.3 %
	#103 lPort:25 tos:0 prot:TCP	6.9 %
	#99 ASremote:8404 lPort:80 tos:0 prot:TCP	9.4 %
	#100, #101 *rIP:81.221.250.10 lIP:130.82.128.1 lPort:53 prot:UDP*	5.4 %
	#100 #101 *rIP:81.221.250.10 lIP:130.82.128.2 lPort:53 prot:UDP*	5.4 %
	flowtype:10 prot:TCP tos:0	5.4 %
	#100 lPort:53 prot:UDP,tos:0	6.8 %
	#99, #104 rIP:85.3.195.113 ASremote:3303 lIP:160.85.36.247 lPort:80 prot:TCP	17.4 %
Signs	**#113** PotOk Onepkt UDP Unreach	12.7 %
	#114 TRWnom UDP Unreach	21.0 %
	#112 TCP Onepkt Unreach	6.3 %

Table 48: Flow and sign item-sets discovered in period August 2008

The flow item-sets displayed in table 48 are the same as in table 46. They also have exactly the same percentage in occurrence. Please refer to the explanations of table 46.

The sign item-sets {**#113**} are also listed in table 46. Item-set {**#114**} displayed in table 48 indicates slow scanning over a whole network to check, which IP addresses are alive and if they run any UDP application or service. Flow item-set {**#100**} describes a lot of failing UDP traffic directed to port 53. This leads to the conclusions, that some clients trying to contact temporarily unavailable DNS servers.

6.6.4. Results February 2008

The results in table 49 do not list the following flow attributes and their values: *durMs, flowtype:2, dPkts, dOctets.* These attributes are not shown because they do provide very small information for the conclusion about the found flow item-sets. The flow item-sets printed in italic font do not display the *ASremote:21494* attribute.

File	Item-sets	Pct
flows_feb 2008	**#99, #106** IIP:195.176.255.135 tos:0 lPort:80 prot:TCP	5.4
	#100 rIP:81.221.252.10 ASremote:21494 lPort:53 prot:UDP	8.2
	#102 IIP:129.132.2.21 tos:0 lPort:123 prot:UDP	5.0
	#100, #105 IIP:195.176.20.204 prot:UDP lPort:53	5.6
	#100, #105 IIP:195.176.20.204 lPort:53 prot:UDP tos:0	5.9
	#103 lPort:25 flowtype:10 prot:TCP	5.3
	#103 lPort:25 tos:0 prot:TCP	6.9
	#99 ASremote:8404 lPort:80 tos:0 prot:TCP	9.4
	#100, #101 *rIP:81.221.250.10 IIP:130.82.128.1 lPort:53 prot:UDP*	5.4
	#100 #101 *rIP:81.221.250.10 IIP:130.82.128.2 lPort:53 prot:UDP*	5.4
	flowtype:10 prot:TCP tos:0	5.4
	#100 lPort:53 prot:UDP,tos:0	6.8
	#99, #104 rIP:85.3.195.113 ASremote:3303 IIP:160.85.36.247 lPort:80 prot:TCP	17.4
signs_feb 2008	**#115** PotOk TCP Unreach	7.4%
	#115 Large TCP Unreach	13.7%
	#116 TRWnom UDP Onepkt Unreach	26.1%
	#115 TCP Onepkt Unreach	23.2%

Table 49: Results from reference interval of class Unreachable

The flow item-sets displayed in table 49 are the same as in table 46. They also have exactly the same percentage in occurrence. Please refer to the explanations of table 46.

The results of the sign item-sets listed in table 49 show, that almost all of the item-sets {**#115**} have the TCP sign set. This result can be caused by the amount of flows not matching to the item-set {**#116**}, which is the most occurring in the peak interval. This item-set also has an additional sign set, the *Onepkt* sign.

The combination of signs in item-set {**#115**} shows a lot of failing TCP connections. This is indicated by having only one packet in the flow, representing the initial ACK. The flow only has one packet, because the responding ICMP message is treated as a new flow.

7. FIM analysis over whole peak periods

7.1. Analysis Setup

When analyzing the whole peak period, this is done by writing the flows to output files covering three hours of the whole period. The output files are named after the first 10 minutes sample of the three hour interval and hold the one way flows of a inspected class.

Processing such a large amount of data takes a long time. Having access to a machine with 16 CPU cores allows it to split the period into 12 intervals, so the data can be processed parallel.

The output files, holding flows and their according sing files, are then analyzed with the FIM tool sam, which produces the item-set files. The analysis is done with a support of 5%. These item-set files are then processed with the `formatCSV` tool to bring them in the CSV format which allows to print graphs over all item-sets inspected in this period. The resulting CSV files do not show the following attributes: *durMs*, *startMs*, *dOctets* and *dPkts*. Item-sets which provide to general informations are not covered in the charts, because they do not allow to find the cause of the traffic.

7.2. Expectations

The results of this analysis are used to verify if the inspected three hour interval is representative for the whole period. It is expected, that the analysis of the whole period shows, that the chosen three hour interval represents not all, but a big portion of the results gathered in the analyzed period.

Another expectation is to find some hints leading to the cause of the peak in the analyzed period. The peak can be caused by flows which didn't appear in the periods before or by flows which where already present, but did not have such a high appearance in the periods without peak.

7.3. Results Class Other Malicious

7.3.1. Flow Item-Sets

Table 50 shows all item-sets, which are newly discovered in the analysis of the whole peak period. Item-sets, like (*prot:UDP,tos:0*), which provide very general informations are not listed in this table.

lPort:53,prot:UDP,tos:0,
lPort:1433,rPort:6000,prot:TCP,ASremote:4134,tos:0
#117 lIP:160.85.182.113,lPort:9,prot:TCP,tos:0
#117 lIP:160.85.182.113,lPort:9,prot:TCP,ASremote:4134,tos:0
rPort:80,prot:TCP,ASremote:17672
rIP:124.238.252.233,rPort:80,prot:TCP,ASremote:17672,tos:8
rPort:80,prot:TCP,ASremote:17672,tos:8
lPort:0,tos:0
lIP:153.109.191.60,prot:UDP,ASremote:4134
lIP:153.109.191.60,prot:UDP,tos:0
rPort:80,prot:TCP,ASremote:4837,tos:0
rPort:80,prot:TCP,tos:0
rPort:80,prot:TCP
rIP:221.10.253.193,rPort:80,prot:TCP,ASremote:4837,tos:0
lIP:160.98.20.58,prot:UDP
rPort:80,prot:TCP,ASremote:16626,tos:0
#116 lPort:1433,rIP:84.244.182.216,prot:TCP,tos:0

Table 50: New item-sets in analysis of whole period

IP and Port in item-set	IP Owner and Port Description
lPort:53,prot:UDP	53/UDP: DNS
lPort:1433,rPort:6000,prot:TCP	1433/TCP: MS SQL Server remote
lIP:160.85.182.113,lPort:9,prot:TCP	9/TCP & UDP: Discard Service; ZHAW
rPort:80,prot:TCP	80/TCP: HTTP
rIP:124.238.252.233,rPort:80,prot:TCP	80/TCP: HTTP; IP address is from China
lPort:0	Port: 0 could indicate ICMP traffic
lIP:153.109.191.60,prot:UDP	Haute Ecole Valaisanne (HEVS)
rIP:221.10.253.193,rPort:80,prot:TCP	80/TCP: HTTP; China Communications
lIP:160.98.20.58,prot:UDP	Ecole d'ingenieurs Fribourg
lPort:1433,rIP:84.244.182.216,prot:TCP	1433/TCP: MS SQL Server remote; IP from provider in NL

Table 51: Explanation of observed ports and IP's

Shown in table 51 are the most important information of the new item-sets. Note that no duplicate entries exist in this table. The informations include source and destination IP and port, as well as the protocol used. The listed item-sets must not provide all of these five attributes. The other column of table 51 hold the information, which application protocol is used. It also provides the information of the owner of the IP address.

The observation of two new protocols in the analysis over the whole period are made. The Microsoft SQL Server [39] uses port 1433/TCP to provide remote access to the database as described in item-set {**#116**}. Item-set {**#117**} uses the Discard Service on port 9 defined in RFC 863 [40]. Table 50 shows this item-set with the TCP protocol, but this service also operates on UDP.

On the next page, illustration 7 shows a chart of all flow item-sets found in this period. The ones providing to general information are not covered. The chart reveals two item-sets with peaks in their occurence. First, the (*lIP:129.132.2.21,lPort:37,prot:UDP,tos:0*) item-set, which has one significant peak in the interval of the 05.08.2008 around 14.10. The other is the (*lIP:129.132.2.21,ASremote:4143,prot:UDP,tos:0*) item-set, which shows significant peaks at the begining and in the middle of the observed period. The last item-set to mention is the (*lIP:129.132.2.21,lPort:37,ASremote:4143,prot:UDP,tos:0*), which also shows a small peak in the same time period.

According to a statistic of flow amounts over this period, there should be a significant peak between August the 1st 2008 0:00 and August the 5th 2008 14:00. This period only shows one item-set present, which is (*lIP:129.132.2.21,lPort:37,ASremote:4143,prot:UDP,tos:0*). At the end of this time period, there is another item-set (*lIP:129.132.2.21,lPort:37,prot:UDP,tos:0*) showing a significant peak.

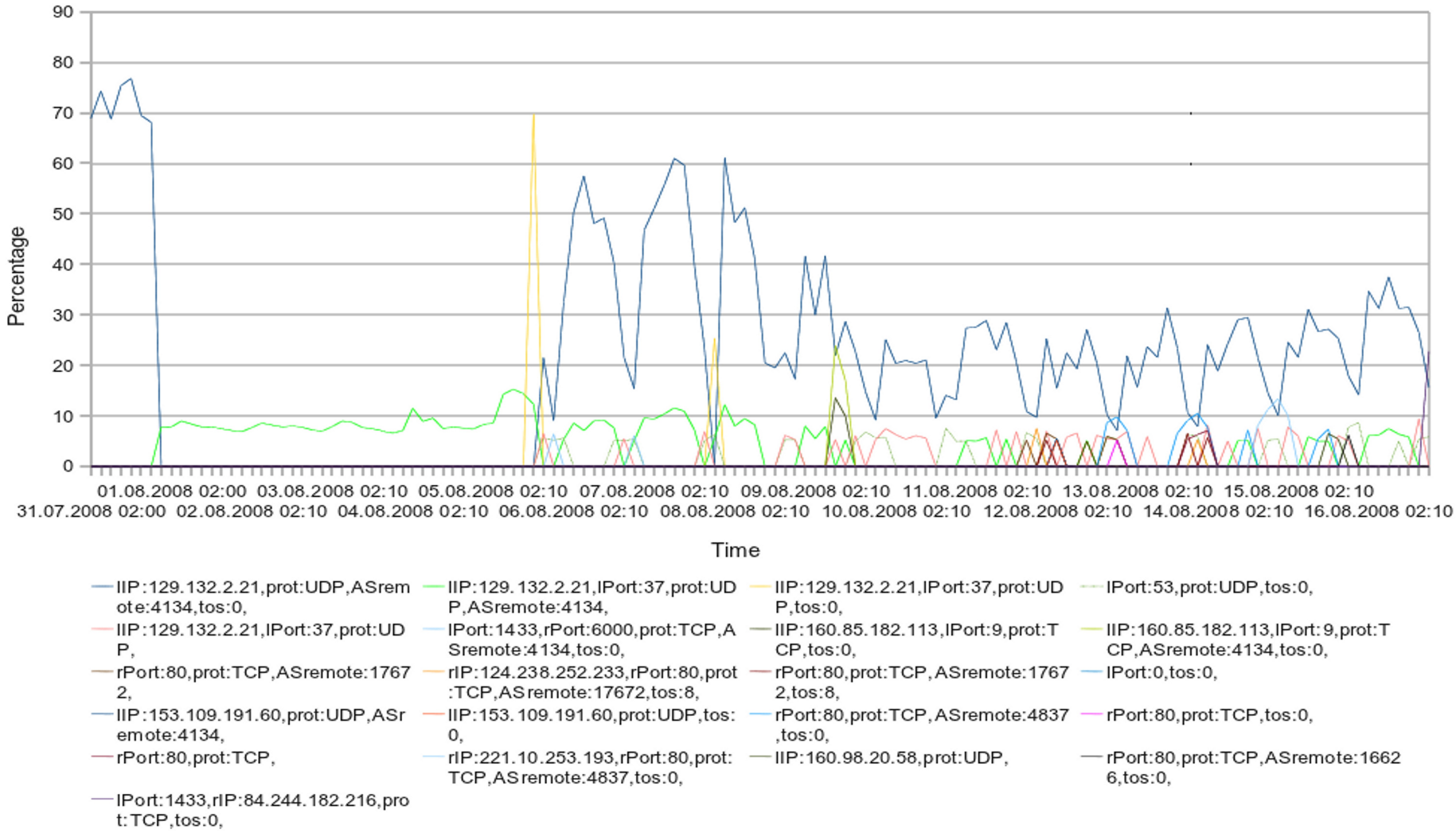

Illustration 7: Results of Flow Item-Sets over whole peak period in August 2008 of class Other Malicious

7.3.2. Sign Item-Sets

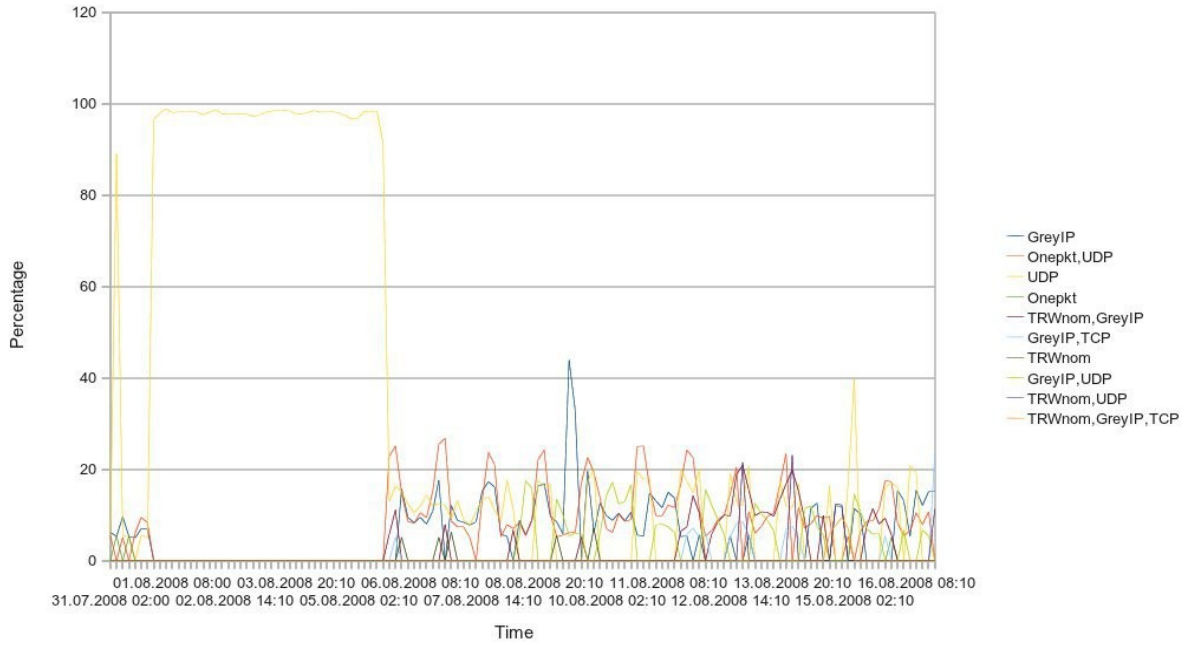

Illustration 8: Sign Item-Sets over whole period of August 2008

The chart shown in illustration 8 the UDP item-set reaches almost 100 % in the first third of the chart. Its portion is reduced, because other item-sets also use the UDP sign. Analyzing the whole interval does bring up some new item-sets, as described in table 52.

UDP
Onepkt
TRWnom, GreyIP
TRWnom
TRWnom, GreyIP, TCP

Table 52: Newly discovered Item-Sets in period August 2008

These sign item-sets are combined out of the item-sets discovered in the analysis over a three hour interval.

7.4. Results Class Backscatter

7.4.1. Flow Item-sets Februar 2007

Analysing the whole period has brought up one item-set not observed in the analysis of a three hour sample interval:

lPort:0,rIP:129.196.226.51,rPort:8,prot:ICMP

Fluke Corporation does own this IP address. The registar information is gathered by ARIN. Following, illustration 9 displays a graph showing all appearing item-sets in this period without reducing them.

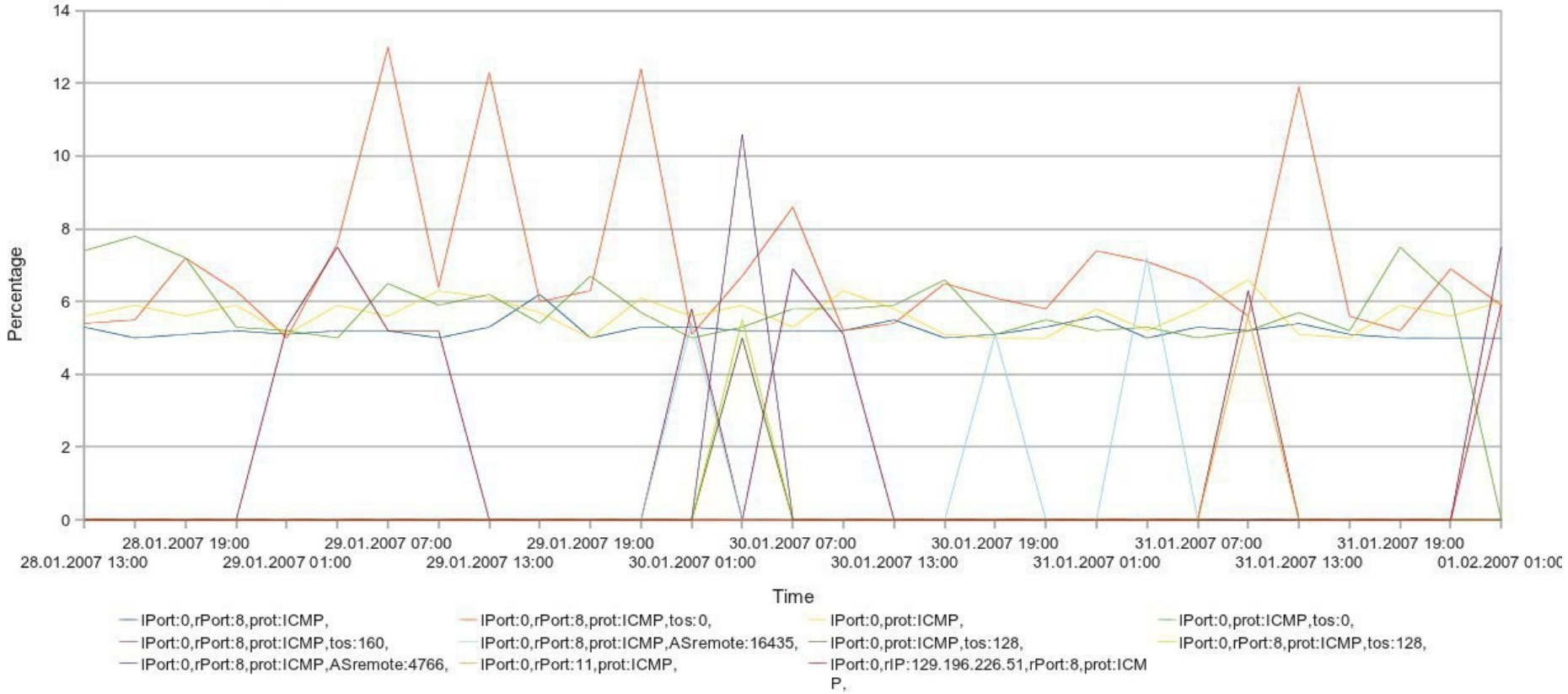

Illustration 9: Chart containing all found Flow Item-Sets of Period Februar 2007 without Reduction of the class Backscatter

7.4.2. Flow Item-sets August 2007

Table 53 contains the newly observed item-sets in the period of August 2007. In addition, the owner of the IP address is listed in this table.

Item-sets	IP Description
lPort:0,rIP:200.57.20.40,rPort:8,prot:ICMP,tos:128	Cable Net International S.A. Mexico (APIC)
lPort:0,rIP:66.206.50.158,rPort:8,prot:ICMP,tos:0	MARQUETTE-ADAMS TEL. COOP. INC (ARIN)
lPort:0,rIP:200.215.83.231,rPort:8,prot:ICMP,tos:0	Comite Gestor da Internet no Brasil (LACNIC)

Table 53: New developed Flow Item-sets in August 2007

For a better overview in the chart, we reduce the item-sets listed in table 54 and add up their percentage for each interval in the whole period. This also includes adding up the item-set *lPort:0,rPort:8,prot:ICMP,tos:0*, which holds all flows only having these attributes, with different ASremote information.

Item-Set	Reduced Item-Set
lPort:0,rPort:8,prot:ICMP,ASremote:xxxx, tos:0	*lPort:0,rPort:8,prot:ICMP,tos:0*
lPort:0,rPort:8,prot:ICMP,ASremote:xxxx, tos:128	*lPort:0,rPort:8,prot:ICMP,tos:128*

Table 54: Reduced Item-Sets in period August 2007

Illustration 10 displaying a chart, containing the reduced item-sets over the whole period. It also contains the item-sets discovered in the FIM analysis with a support of 5%.

As it can be observed, the item-set (*lPort:0, rPort:8, prot:ICMP, tos:0*) shows some significant peaks int this period. In twice peaks, it covers almost covers 40 % of the analyzed flows of this period.

The second item-set, which only shows twice significant peaks, is the item-set (*lPort:0, rPort:8, prot:ICMP, tos:128*).

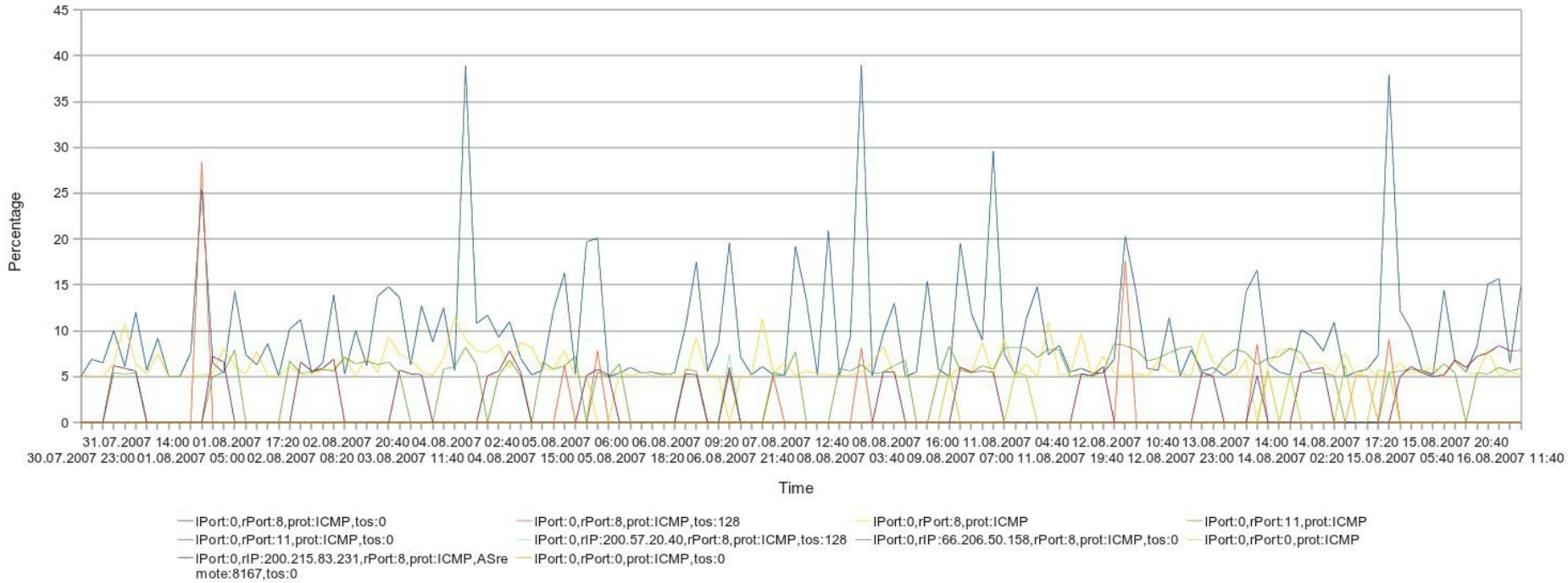

Illustration 10: Chart showing the reduced Flow Item-Sets of Period August 2007 of the class Backscatter

7.4.3. Flow Item-sets Februar 2008

Table 55 contains all new discovered item-sets in the analysis of the whole period of the class Backscatter in February 2008. Note that the *tos:0* attribute is removed by all item-sets for displaying them on one line. The second column of the table contains the owner of the IP and the registrar providing the informations.

Item-Sets	**IP Description**
lPort:0,rIP:129.196.226.36,rPort:8,prot:ICMP	Fluke Corporation (ARIN)
lPort:0,rIP:221.130.180.223,rPort:771,prot:ICMP	China mobile communications (RIPE)
lPort:0,rIP:129.196.226.21,rPort:8,prot:ICMP	Fluke Corporation (ARIN)
lPort:0,rIP:131.109.100.22,rPort:8,prot:ICMP	Rhode Island Network for Educ. Techn. (ARIN)
lPort:0,rIP:24.144.45.242,rPort:8,prot:ICMP	Conway Coorperation (ARIN)

Table 55: Listing of all new observed item-sets in period February 2008

To reduce the number of item-sets observed in the whole interval, the *ASremote* attribute in the item-sets is removed. This reduces the complexity of the graph and makes it easier to read. Table 56 shows the original and reduced item-sets.

Item-Set	**Reduced Item-Set**
lPort:0,rPort:8,prot:ICMP,ASremote:xxxx, tos:0	*lPort:0,rPort:8,prot:ICMP,tos:0*
lPort:0,rPort:8,prot:ICMP,ASremote:xxxx, tos:128	*lPort:0,rPort:8,prot:ICMP,tos:128*

Table 56: Listing of original and reduced item-sets in February 2008

Illustration 11 shows all item-sets found over the whole period of February 2008, except the ones, which provide to general informations.

The period covered in illustration 11 shows one item-set having a lot of peaks over 40 %. It is the same item-set, which also has the highest peaks in illustration 10. The item-set contains the items (*lport:0, rPort:8, prot:ICMP, tos:0*).

The other item-set with peaks in illustration 10 only shows one significant peak in illustration 11. This is the item-set (*lPort:0, rPort:8, prot:ICMP, tos:128*).

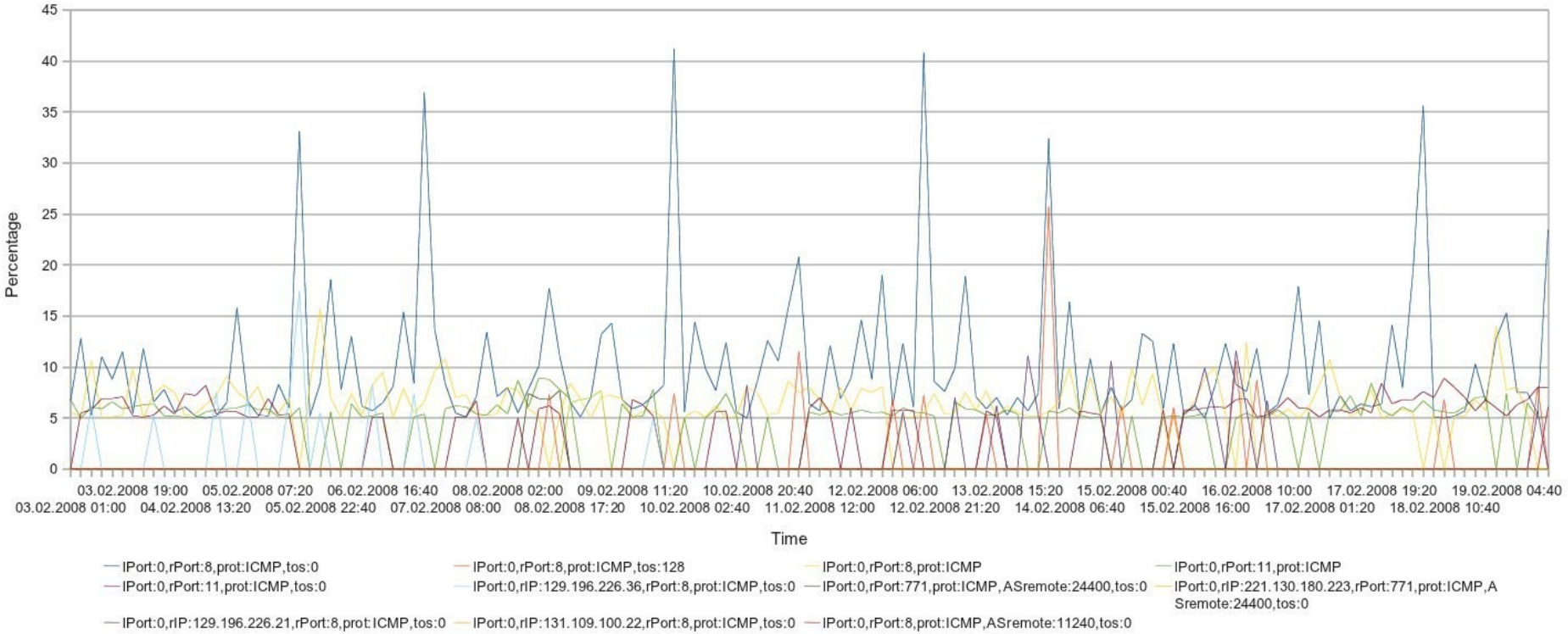

Illustration 11: Chart displaying all found Flow Item-Sets over whole Period of Februar 2008

7.4.4. Sign Item-Sets Februar 2007

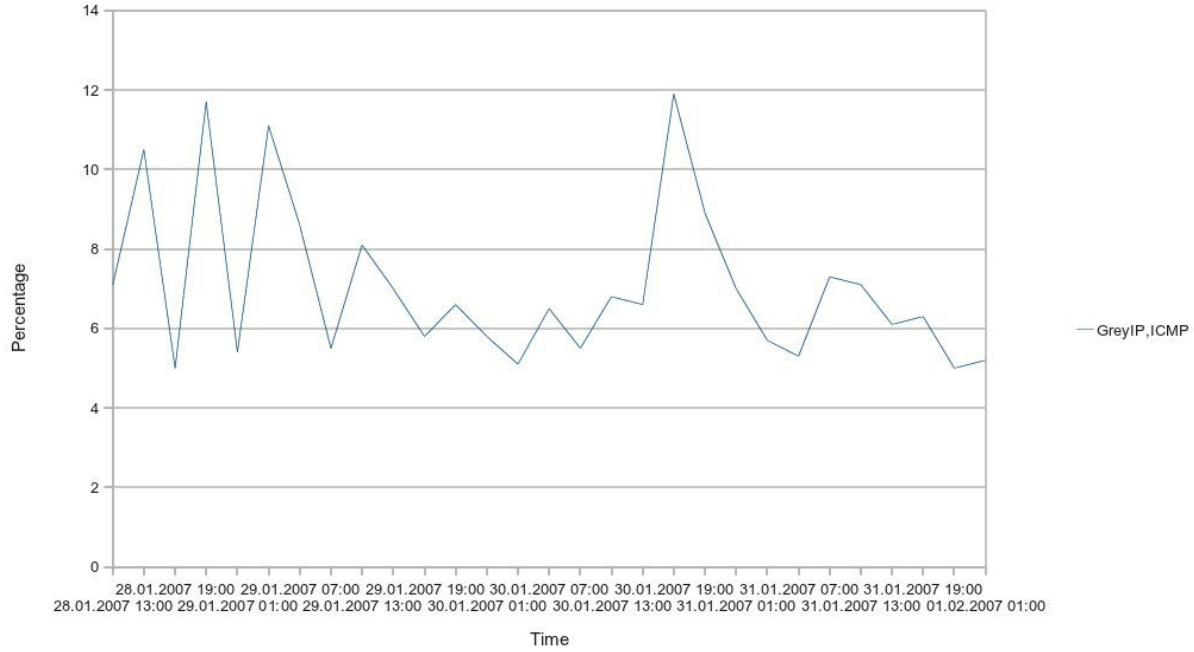

Illustration 12: Chart of discovered Sign Item-Sets in Period Februar 2007

Illustration 12 shows the percentage of the only item-set discovered in the analysis of the whole period. The analysis is made with a support of 5 %, but the item-set observed over the whole period is a more general one. It is not splited up into two item-sets, as it can be observed in the analysis of a continous three hour interval. The splited up item-sets were (*Onepkt*, *GreyIP*, *ICMP*) and (*TRWnom*, *GreyIP*, *ICMP*).

7.4.5. Sign Item-Sets August 2007

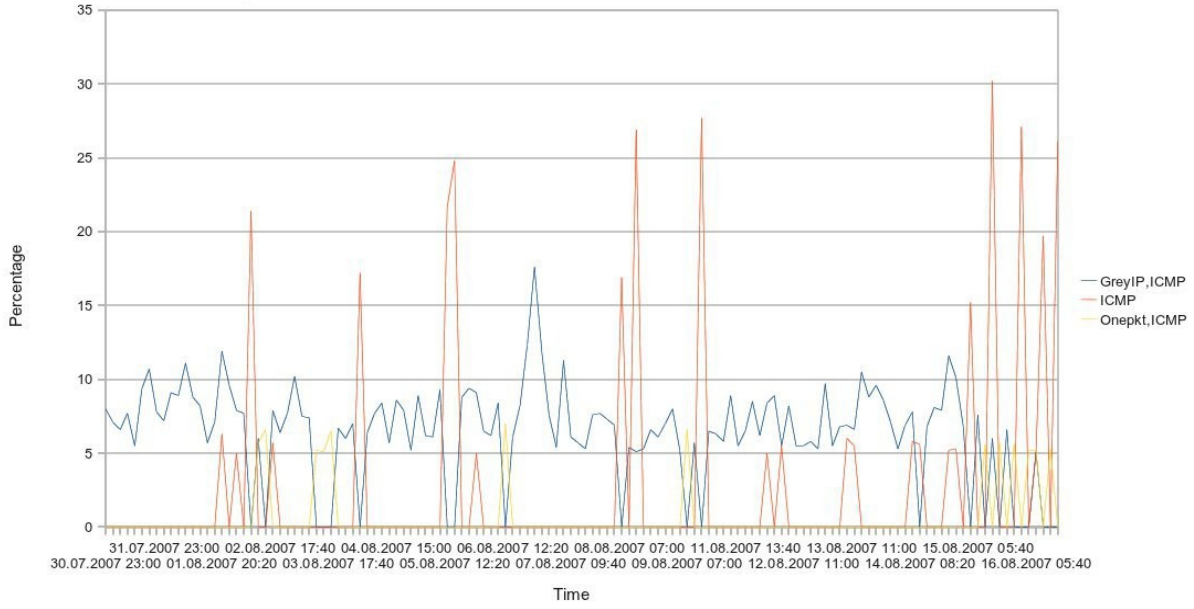

Illustration 13: Discovered Sign Item-Sets in Period August 2007

All item-sets found in this period have the *ICMP* sign as shown in illustration 13. The more general ICMP item-set is first found in the analysis of the whole period. Also new are the two item-sets

(*Onepkt*, *ICMP*) and (*GreyIP*, *ICMP*). These are a split up from the more general (*GreyIP*, *Onepkt*, *ICMP*) item-set, discovered in the analysis of a continous three hour interval of this period.

7.4.6. Sign Item-Sets Februar 2008

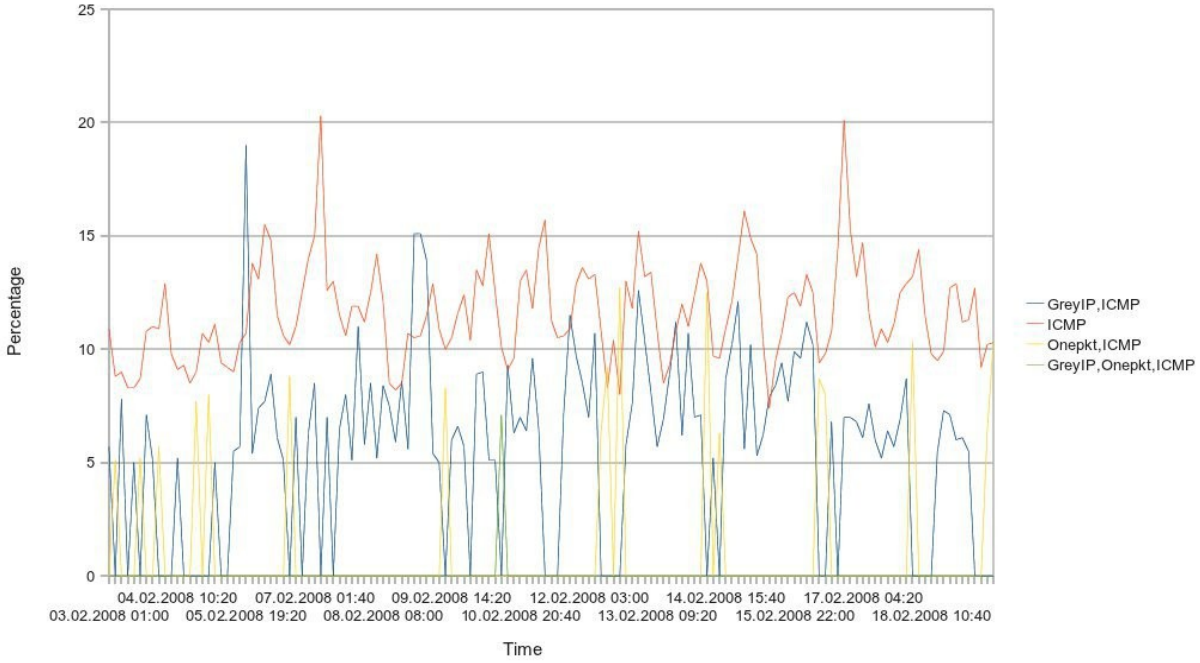

Illustration 14: Chart covering all found Sign Item-Sets in Period Februar 2008

Looking at the results of the analysis over the whole period of Februar 2008 in illustration 14, it can be observed that the (*Large*, *ICMP*) item-set is missing, which shows up in the analysis of a continous three hour interval of this period.

The item-set having only the *ICMP* sign set could cover the (*Large*, *ICMP*) flows due to its appearance of more than 10 % in averange. Another new item-set in the analysis of the whole period is the (*GreyIP*, *Onepkt*, *ICMP*) item-set, which combines the (*GreyIP*, *ICMP*) and (*Onepkt*, *ICMP*) item-sets. The occurence of the new item-set is small, it appears only within a sporadicaly peak.

7.5. Results Class Benign P2P

7.5.1. Flow Item-sets August 2007

Table 57 shows all new item-sets without the ones allready discovered in the FIM analysis with a support of 5%.

lPort:11000,prot:UDP,tos:0,
lIP:130.92.70.252,lPort:11000,prot:UDP,tos:0,
lIP:129.194.58.16,prot:UDP,
lIP:129.194.58.16,lPort:1262,rPort:6346,prot:UDP,tos:0,
lIP:129.194.58.16,lPort:1262,prot:UDP,tos:0,
lIP:147.86.200.5,
lIP:129.194.58.16,lPort:1262,prot:UDP,
lIP:129.194.58.16,prot:UDP,tos:0,
lIP:147.86.200.5,prot:TCP,
lIP:147.86.200.5,prot:TCP,tos:0,
lPort:11000,prot:UDP,ASremote:11318,tos:0,
lIP:192.41.135.219,lPort:11000,prot:UDP,ASremote:87,tos:0,
lIP:192.41.135.218,lPort:11000,rIP:193.1.201.27,prot:UDP,ASremote:1213,tos:0,
lIP:129.194.58.16,lPort:1829,prot:UDP,tos:0,
lIP:129.194.58.16,lPort:1829,prot:UDP,
lIP:129.194.58.16,lPort:1829,rPort:6346,prot:UDP,tos:0,
lIP:147.86.200.5,tos:0,
lPort:11000,prot:UDP,
lIP:130.92.38.110,prot:UDP,
lIP:129.132.131.65,prot:UDP,tos:0,
lIP:129.132.31.209,prot:UDP,tos:0,
lIP:195.176.54.160,prot:UDP,tos:0,
lIP:82.130.65.196,prot:UDP,tos:0,
lIP:129.132.131.65,prot:UDP,
lIP:129.132.130.199,prot:UDP,tos:0,
lIP:129.132.59.48,prot:UDP,tos:0,
lIP:129.132.108.61,prot:UDP,tos:0,
lIP:82.130.120.66,prot:UDP,tos:0,
lIP:130.60.204.16,prot:UDP,
lIP:130.60.204.16,prot:UDP,tos:0,
lIP:129.132.177.216,prot:UDP,tos:0,
lIP:129.132.23.194,prot:UDP,tos:0,

Table 57: Listing of all new item-sets discovered in analysis of whole period in August 2007

The explanation of the owner of the IP address can be extracten from table 58. It also provides the information, about the listed ports of table 57. The IP addresses and port informations are only displayed once, eliminating the redundancy.

Port:1262/UDP	QNTS-ORB
Port:1829/UDP	Optika eMedia
Port:6346/UDP	gnutella
Port:11000/UDP	Various applications
IP:82.130.64.0/18	ETH
IP:129.132.0.0/16	ETH
IP:129.194.58.16	University of Geneva
IP:130.60.204.16	University of Zurich
IP:130.92.0.0/16	University of Bern
IP:147.86.200.5	Fachhochschule Nordwest Schweiz
IP:192.41.135.0/24	University of Zurich
IP:193.1.201.27	Planet Lab Network
IP:195.176.54.160	Universita della Svizzera italiana

Table 58: Listing explaining the IP addresses and Ports found in Period August 2007

Port 1829/UDP is used by the Optika eMedia service. This service is part of the Oracle Imaging and Process Management [41].

The port 11000/UDP is used by various applications and two worms [42]. By default IANA has assigned the IRISA service to it. Other applications using this port are: Everquest Online Adventures, Cisco Border Gateway Protocol, Microsoft Visual Studio, .Net Framework, SCInterface, The Matrix Online (TCP) and Archlord. Malicious traffic on this port is caused by the Senna Spy Trojan Generator [43].

It could not be found out, what the QNTS service is. It is using port 1262/UDP. The ORB extension could indicate a Object Request Broker, as normally found in CORBA.

The chart in illustration 15 does not reduce any item-sets, because the results have a lot of different IP addresses in it.

The chart shown in illustration 15 shows a begining peak at the end of the observated interval. An item-set can observed, which shows a peak. This is the item-set (*IIP:82.130.65.196,prot:UDP,tos:0*). It starts raising at the end of the inspected period.

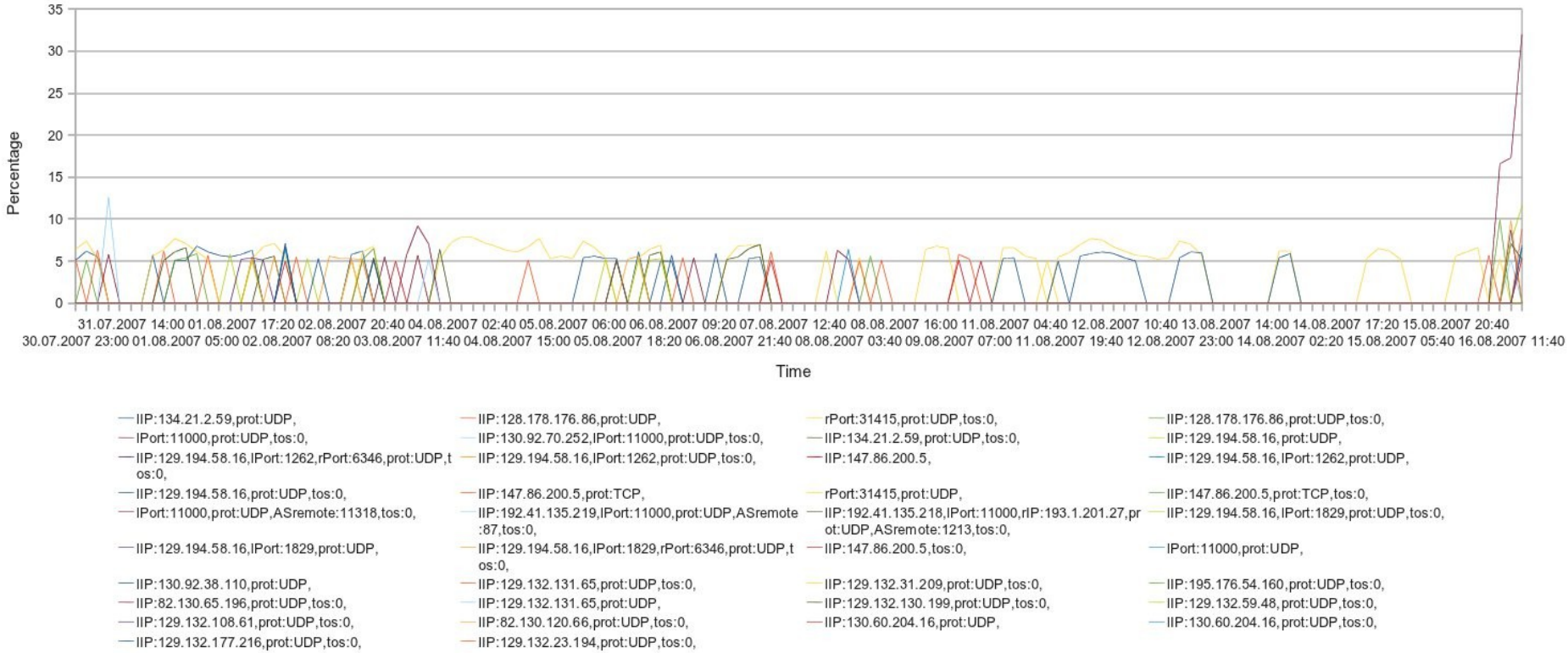

Illustration 15: Discovered Flow Item-Sets in Period August 2008 without Reduction

7.5.2. Flow Item-sets Februar 2008

The found item-sets in the whole period of Februar 2008 in the class unreachable does not bring up any new item-sets. All informative item-sets found list the local IP's 82.130.102.161 or 82.130.102.218 or both, referencing the fake e-mule server. Also the UDP ports 4246 and 4254 show up, which had been assigned to the fake e-mule servers.

For a better overview in the graph, we reduce the item-sets listed in table 59 to more general item-sets.

Item-Sets	**Reduced Item-Sets**
Lport:4246,prot:UDP	*Lport:4246,prot:UDP*
lPort:4246,prot:UDP,tos:0	
IIP:82.130.102.218,lPort:4246,prot:UDP,tos:0	*IIP:82.130.102.218,lPort:4246,prot:UDP*
IIP:82.130.102.218,lPort:4246,prot:UDP	
IIP:82.130.102.161,lPort:4246,prot:UDP,tos:0	*IIP:82.130.102.161,lPort:4246,prot:UDP*
IIP:82.130.102.161,lPort:4246,prot:UDP	
IIP:82.130.102.161,prot:UDP	*IIP:82.130.102.161,prot:UDP*
IIP:82.130.102.161,prot:UDP,tos:0	
IIP:82.130.102.218,prot:UDP	*IIP:82.130.102.218,prot:UDP*
IIP:82.130.102.218,prot:UDP,tos:0	

Table 59: Item-Sets and their resulting Reduction in Period Februar 2008

Illustration 16 on the next side shows the resulting chart, when reducing the found item-sets.

The item-sets shown in illustration 16 do not have any significant peaks. The all have a percentage of flows which is under 16 %. This is not much when considering, that the item-sets shown are reduced item-sets, which have counted up percentages. So the real item-sets would have much lower percentage of matching flows.

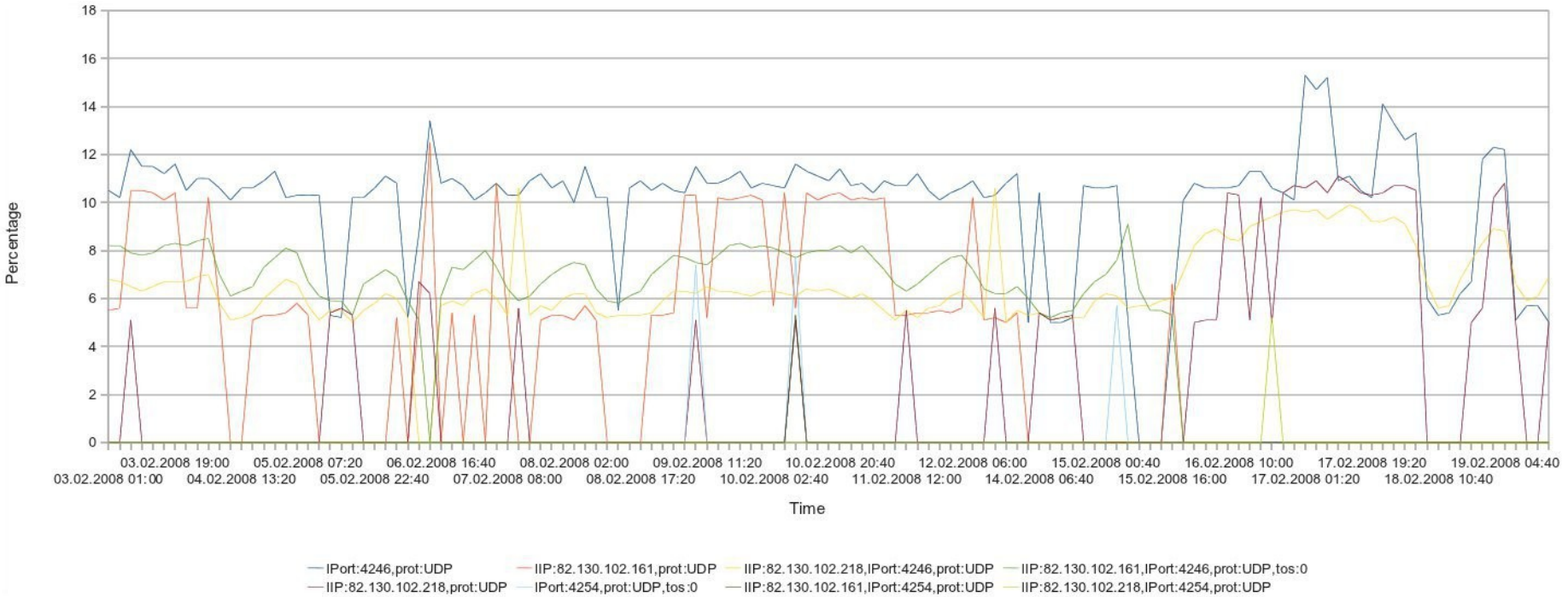

Illustration 16: Observed Flow Item-Sets in Period Februar 2008

7.5.3. Sign Item-Sets August 2007

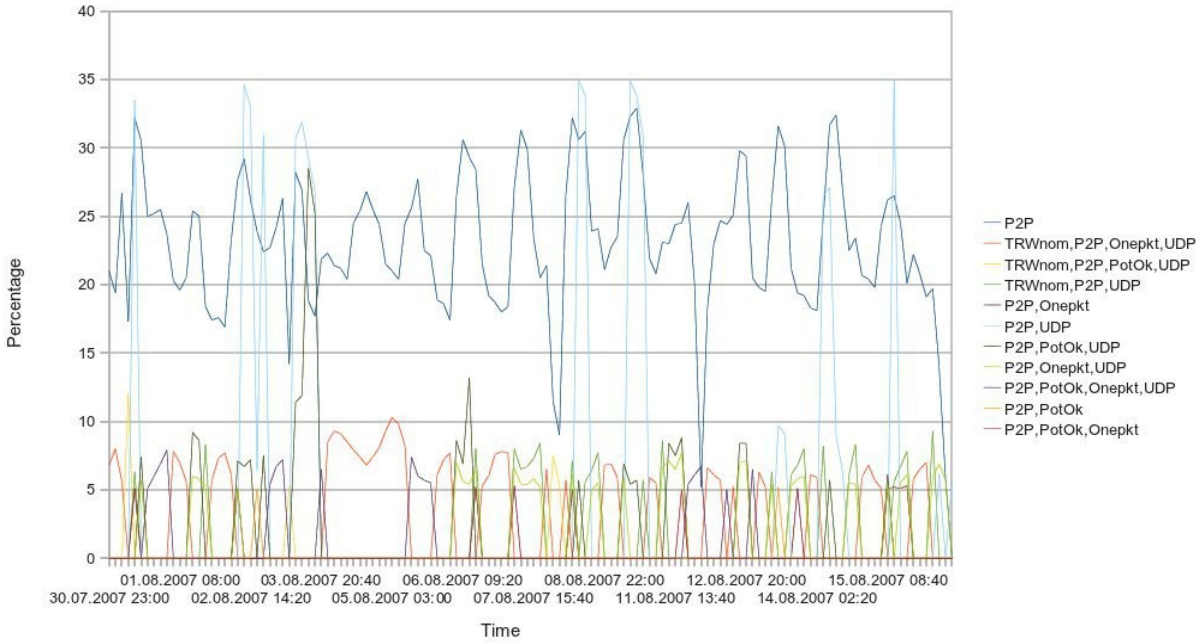

Illustration 17: Sign Item-Sets observed in Period August 2007

In the results of the whole period shown in illustration 17, no item-set can be discovered having the *TCP* sign set. The (*PotOk*, *TRWnom*, *Onepkt*, *P2P*, *UDP*) item-set found in the analysis of a continous three hour interval, is splited up into more, finer grained item-sets.

7.5.4. Sign Item-Sets Februar 2008

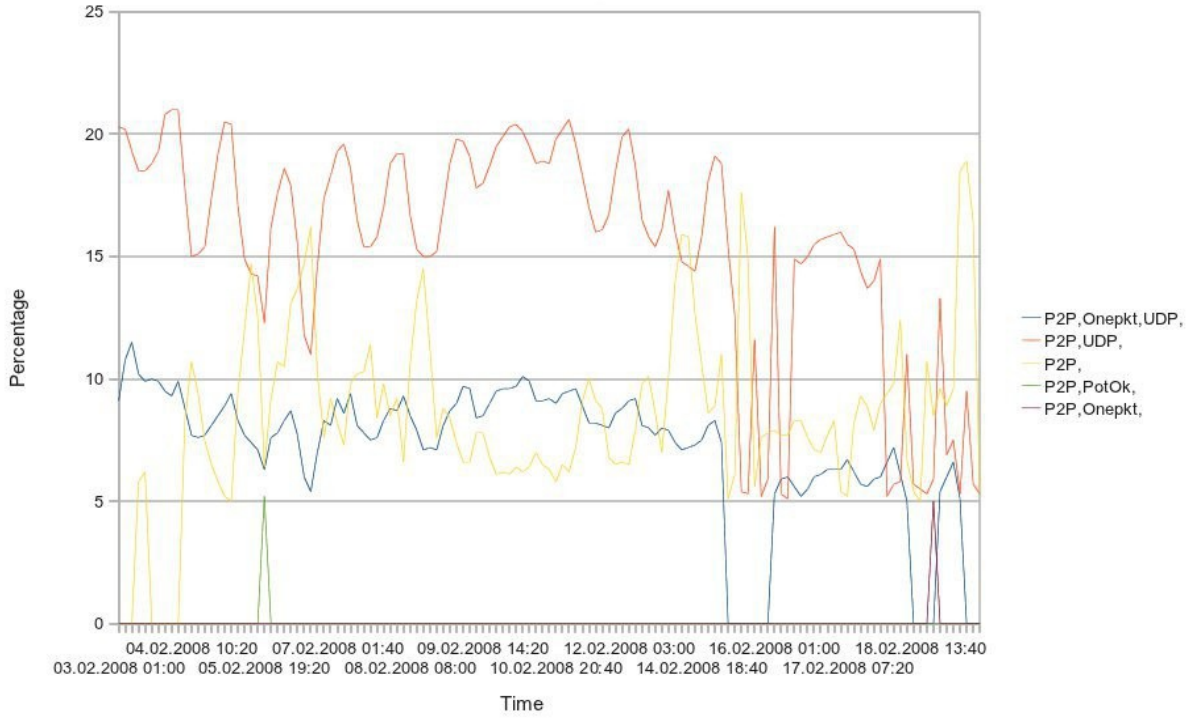

Illustration 18: Sign Item-Sets observed in Period Februar 2008

Illustration 18 shows the found sign item-sets in this period. The item-set (*Large*, *UDP*, *P2P*) doesn't show up in the analysis over the whole period. It can also be observed, that the (*PotOk*, *Onepkt*, *UDP*, *s*) item-set is splited up into finer grained item-sets.

7.6. Results Class Unreachable

7.6.1. Flow Item-Sets August 2008

Table 60 shows all item-sets which are found additionaly in the analysis over the whole period. Item-sets allready discovered in

Item-Sets	Port / IP Description
lPort:53,prot:UDP,tos:0	Port 53/UDP: DNS lookups
lIP:129.195.254.33,prot:TCP,tos:0	University of Geneva
lPort:53,prot:UDP	Port 53/UDP: DNS Lookups
lPort:53,tos:0	Port 53/UDP, TCP: DNS lookup or zone transfers

Table 60: Newly observed Flow Item-Sets in August 2008

Illustration 19 shows the whole period with all item-sets providing specific information of flows.

A significant peak of a item-set shown in illustration 19 occurres almost at the end of the chart. This item-set describes normally benign traffic, which is directed to a NTP server on port 123/UDP.

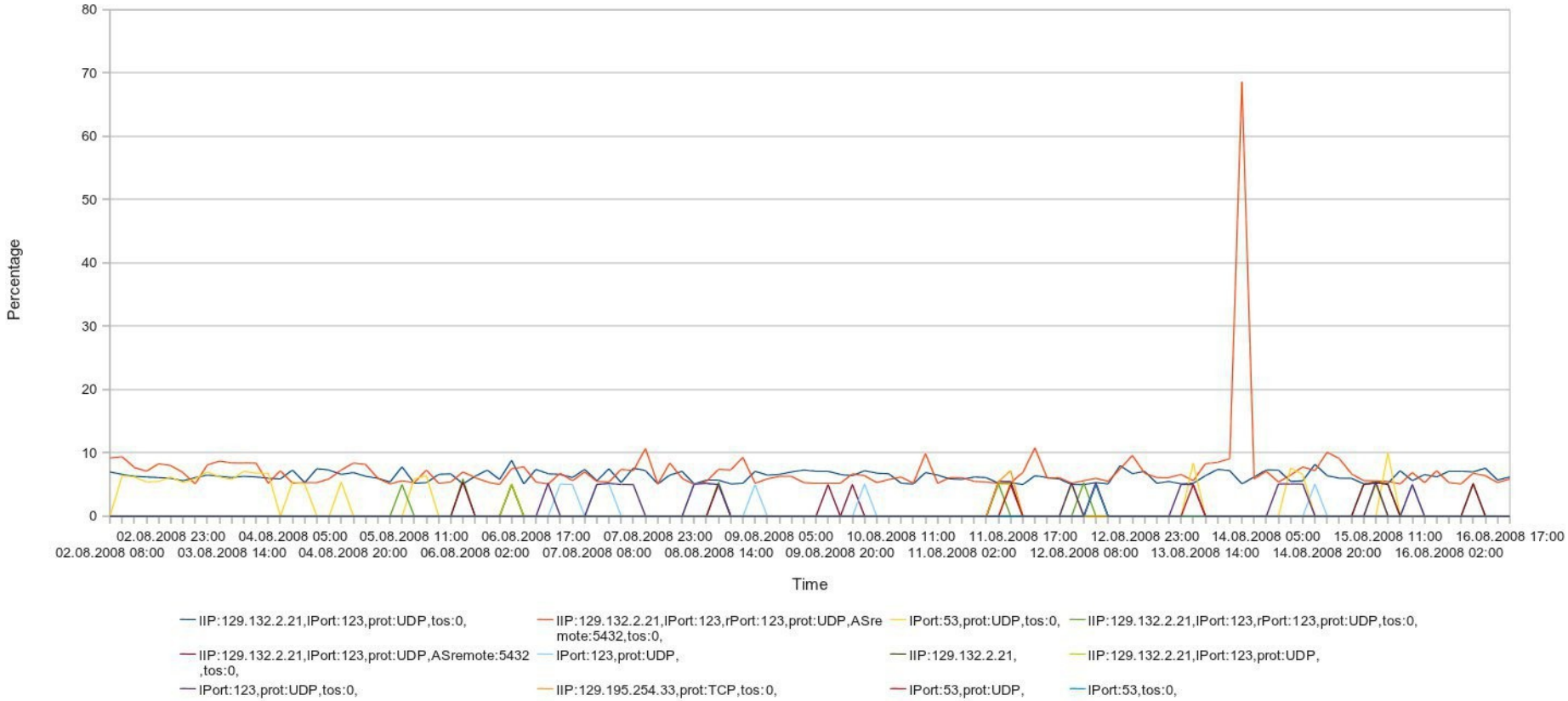

Illustration 19: Flow Item-Sets over whole Period August 2008

7.6.2. Flow Item-Sets Februar 2009

Table 61 lists all new discovered item-sets in the analysis of the whole period. The IP shown in the second item-set belongs to SWITCH and has something to do with the MERAPI, a JAVA air message bridge.

Item-Sets	**Port and IP Description**
lPort:53,prot:UDP,tos:0	Port 53/UDP: DNS lookup
lIP:130.59.211.10,lPort:53,prot:UDP,tos:0	Port 53/UDP: DNS lookup, IP: SWITCH

Table 61: Newly discovered Item-Sets in Period Februar 2009 with IP owner

On the next page, illustration 20 shows a chart over all item-sets found in this period. The chart displays all found item-sets. None of them are reduced. Only item-sets, which provide to general information and do not help to identify which one way flows they are describing, are filtered.

The shown flow item-sets displayed in illustration 20 do not show any anomallies or peaks. Most of the flows applying to these item-sets are benign traffic directed to NTP or DNS servers which is indicated by the local port used.

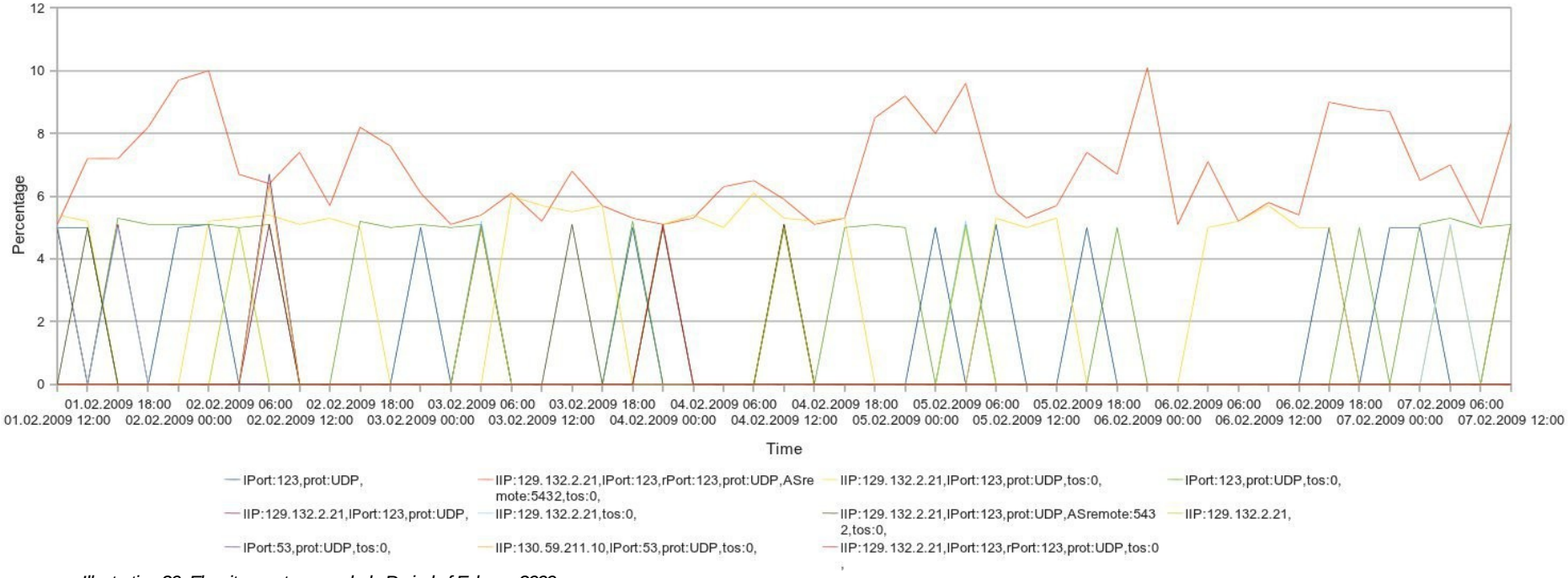

Illustration 20: Flow item-sets over whole Period of Februar 2009

7.6.3. Sign Item-Sets August 2008

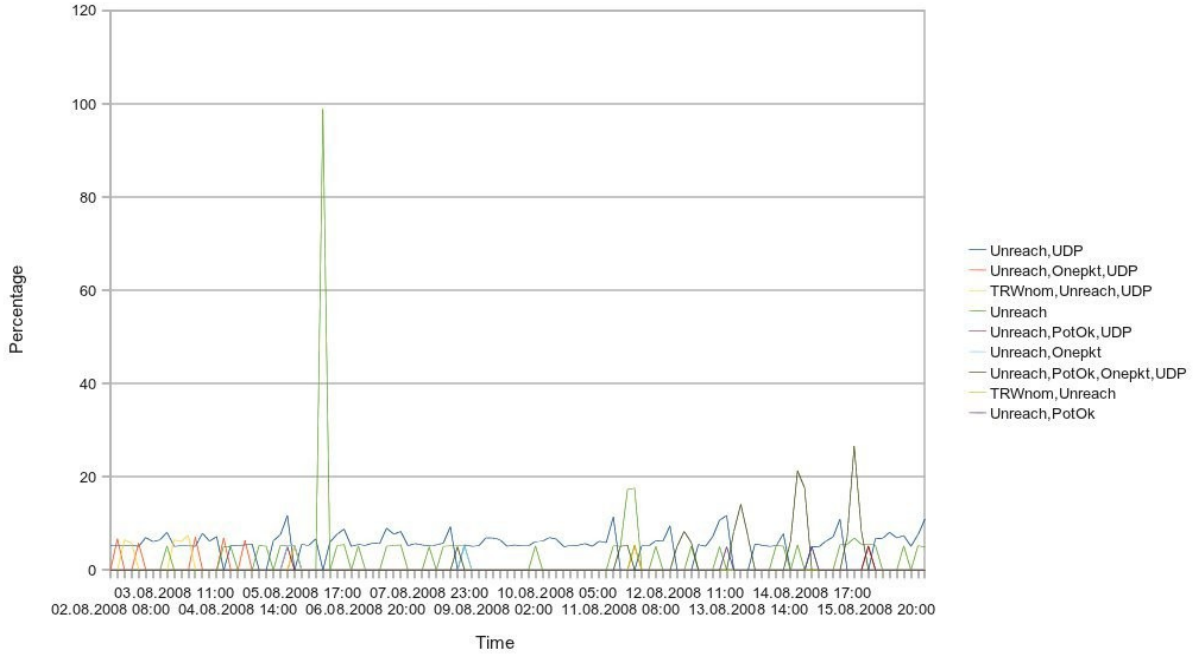

Illustration 21: Chart showing the found Item-Set in Period August 2008

Analyzing the whole period brings up some new sign item-sets as shown in illustration 21. It can be observed, that the new item-sets consist out of the same signs as the ones found in the analysis of a three hour interval. One item-set shows a significant peak, which is the single signed Unreach item-set. This is not suspecting because there is only one rule to assign flows to this class which uses the Unreach sign.

7.6.4. Sign Item-Sets February 2009

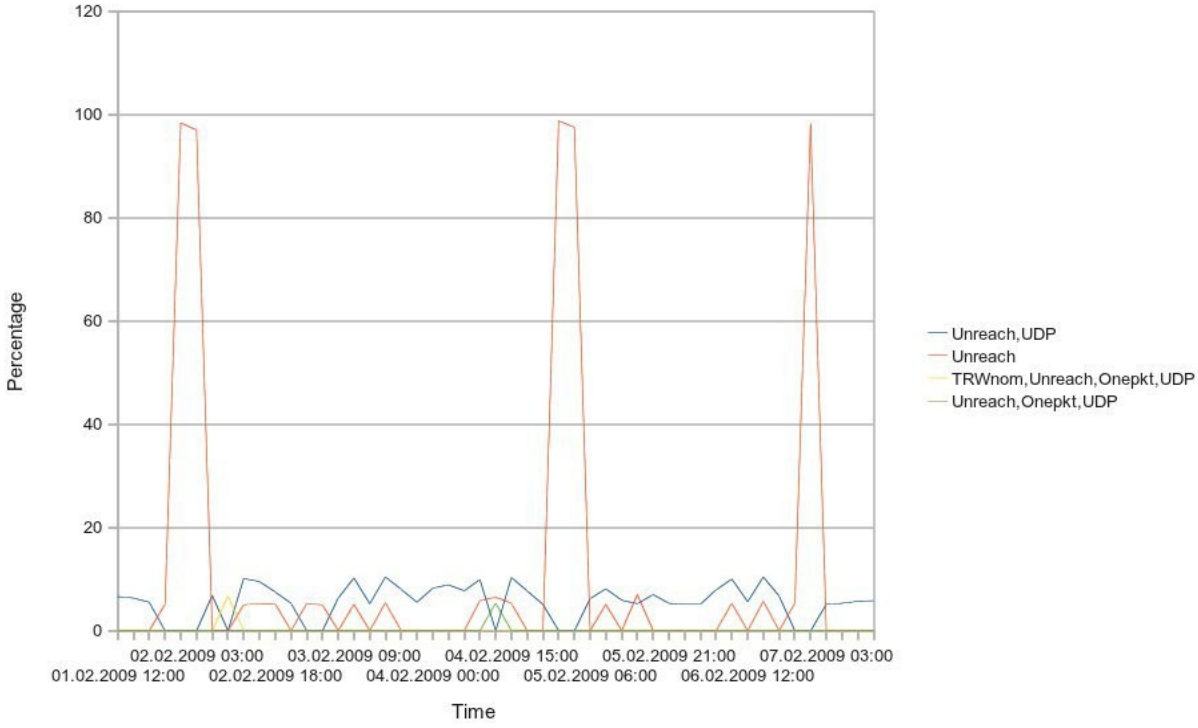

Illustration 22: Chart covering all Sing Item-Sets found in Period Februar 2009

The chart in illustration 22 does not reveal much information. Only generalized item-sets of the item-set (*Onepkt*, *UDP*, *Unreach*) observed in the analysis of a three hour interval are shown.

8. Sign Occurrence Analysis

8.1. Analysis Setup

To know which signs occur in an interval of a inspected peak period, a statistical analysis, which counts all signs belonging to the interval, is made. The analysis is done by a tool called `classStats` and was developed in the scope of this diploma thesis.

The `correlate_flow_signs` tool counts all signs of flows applying to a flow item-set found in the same interval. Therefore it needs a list of flow item-sets which it checks against the flows of the inspected interval. If the flow matches the item-set, the according line in the sign file is read and the occurrence of signs is being counted.

The output of these two small tools is written to the command line. Therefore it's output should be redirected to a file.

This analysis does only cover item-sets, which are informative. Item-sets which are describing very general information are not considered in this analysis. The analysis is done with the item-sets gained from the FIM analysis with a support of 10%. Only the three hour intervals of the peak periods are inspected. The location of the resulting statistic files are listed in the appendix.

8.2. Expectations

It is expected, that the creation of a statistic per item-set helps to identify, if the item-set is classified correctly. The counting of sing occurrences per class should lead to the discovery of sign characteristics of the analyzed class. This analysis also covers the information, which rules are assigning the most traffic to the classes “Backscatter” and “Other Malicious”.

8.3. Class Other Malicious

8.3.1. Over all Sign Statistics

Period	August 2008			
Total Flow Count	3'488'032			
Counted Signs	Onepkt:	3'097'355	P2P:	33'341
	UDP:	3'232'048	Large:	49'075
	TCP:	255'984	Retry:	195
	TRWnom:	221'306	Artef:	47
	GreyIP:	390'677		

Table 62: Statistical count of all signs occurring in inspected interval of class Other Malicious

8.3.2. Conclusions over all Sign Statistics

The over all sign statistics listed in table 62 of the class „Other Maliciouis“ shows that 92,7% of assigned one way flows are using the UDP protocol. It can also be observed, that a small amount of *TRWnom* signs are present, which could indicate slow or stealth scanning attempts. Another observation is that 8,5 % of traffic assigned by the *GreyIP* rule has the sign *P2P* set. Flows having these signs could indicate failing connection attempts between P2P hosts. This is especially the case, when using outdated peer lists and one IP (the one which leads to the GreyIP sign) is down.

Having a look at the two rules for the class „Other Malicious“ shows the effectivness of these rules. Table 63 shows, that the rule, which assigns most of the flows to this class, is the one with the Onepkt sign.

Period	Onepkt	GreyIP
August 2008	88,8 %	11,2 %

Table 63: Rule effectivenes of Class Other Malicious in August 2008

8.3.3. Signs per Item-set

Total Flow Count	3'488'032			
Item-set	IIP:129.132.2.21 lPort:37 ASremote:4134 tos:0 dOctets:46 dPkts:1 durMs:0 prot:UDP flowtype:2 ASlocal:559			
Flows matching item-set	2'480'457			
Counted Signs	Onepkt: UDP:	2'480'457 2'480'457	TRWnom:	552

Table 64: Statistical count of all signs belonging to flows matching the listed item-set

8.3.4. Conclusions of signs per Item-set

Table 64 shows, that all flows matching to the item-set have the *Onepkt* sign, which indicates failed connection attempts. Only a small amount of the sign *TRWnom* indicates stealth scanning activities. The only question to answer is, why a lot of traffic is originating an AS located in Canada.

8.4. Class Backscatter

8.4.1. Over all Sign Statistics

Period	February 2007			
Total Flow Count	4'927'045			
Counted Signs	Onepkt: ICMP: Large: GreyIP: TRWnom:	1'181'429 4'917'738 85'313 2'054'232 297'206	Backsc: TCP: UDP: Retry: Bogon:	9'771 1'034 8'273 1 1'737

Table 65: Statistical count of all occurring signs in interval February 2007

Period	August 2007			
Total Flow Count	2'560'936			
Counted Signs	Large: ICMP: Onepkt: GreyIP:	101'440 2'558'976 481'977 1'244'992	Backsc: TCP: UDP: TRWnom:	2'159 1'355 605 104'137

Table 66: Statistical count of all occurring signs in interval August 2007

Period	February 2008			
Total Flow Count	3'031'760			
Counted Signs	GreyIP:	749'610	TRWnom:	113'393
	Large:	302'194	Backsc:	1'696
	ICMP:	3'030'293	TCP:	895
	Onepkt:	849'026	UDP:	572

Table 67: Statistical count of all occurring signs in interval February 2008

8.4.2. Conclusions of Interval Sign statistics

The statistics shown in tables 64 to 67 show a small amount of flows using the protocols TCP and UDP. These protocols can only appear in the rule with the *Backsc* sign.

This class has three rules which match traffic to it. Two of them need the *ICMP* sign set, so a distinction between them is needed. The weight of the rule without the *GreyIP* sign is calculated, when the occurrences of the rule with the *GreyIP* sign are subtracted from its occurrences. Table 68 shows the effectiveness in percents of each rule.

Period	**Backsc**	**ICMP**	**GreyIP**
Februar 2007	0,2 %	58,1 %	41,7 %
August 2007	> 0,1 %	51,4 %	48,6 %
Februar 2008	> 0,1 %	75,3 %	24,7 %

Table 68: Rule effectivness of the Class Backscatter

The results of table 68 shows that the rule containing the *Backsc* sign is not very efficient. It only matches a small amount of flows to this class. The rule having only the *ICMP* sign in it, without the *GreyIP* sign, is the most effective. ICMP flows assigned through this rule could also be benign, even if they don't have the *PotOk* sign set.

8.4.3. Signs per Item-set February 2007

Period	February 2007
Total Flow Count	4'927'044

Item-set	rPort:11 lPort:0 prot:ICMP			
Flows matching item-set	616'363			
Counted Signs	ICMP: Onepkt: Large:	616'363 229'419 31'327	GreyIP: TRWnom:	23'005 4

Item-set	dOctets:244 dPkts:4 tos:0 rPort:8 lPort:0 prot:ICMP			
Flows matching item-set	612'776			
Counted Signs	ICMP: TRWnom:	612'776 1'376	GreyIP:	136'030

Item-set	tos:128 lPort:0 prot:ICMP			
Flows matching item-set	720'760			
Counted Signs	ICMP: Large: Onepkt:	720'760 12'694 241'933	GreyIP: TRWnom:	313'805 29'233

Item-set	dPkts:1 durMs:0 rPort:8 lPort:0 prot:ICMP			
Flows matching item-set	657531			
Counted Signs	Onepkt: ICMP:	657'531 657'531	GreyIP: TRWnom:	375'465 19'153

Item-set	dOctets:122 dPkts:2 tos:0 rPort:8 lPort:0 prot:ICMP			
Flows matching item-set	1707102			
Counted Signs	ICMP: GreyIP:	1'707'102 968'483	TRWnom:	243'844

Table 69: Listing of all sign counts belonging to described flow item-sets in Febraury 2007

8.4.4. Conclusion

The sign counts displayed in table 69 are not very informative. It only can be observed, that none of the flow item-sets inspected shows the *Backsc* sign. The rule having only the *ICMP* sign in it is the most effective, followed by the *GreyIP* rule, as listed in table 68.

8.4.5. Signs per Item-set August 2007

Period	August 2007			
Total Flow Count	2'560'935			

Item-set	tos:128 rPort:8 prot:ICMP			
Flows matching item-set	279'783			
Counted Signs	Large: ICMP: GreyIP:	4'647 279'783 139'589	Onepkt: TRWnom:	41'234 45

Item-set	dOctets:244 dPkts:4 tos:0 rPort:8 prot:ICMP			
Flows matching item-set	342'648			
Counted Signs	ICMP:	342'648	GreyIP:	111'674

Item-set	dPkts:1 durMs:0 rPort:8 prot:ICMP			
Flows matching item-set	271'524			
Counted Signs	Onepkt: ICMP:	271'524 271'524	GreyIP: TRWnom:	148'615 5142

Item-set	dOctets:122 dPkts:2 rPort:8 tos:0 prot:ICMP			
Flows matching item-set	1'014'174			
Counted Signs	ICMP: GreyIP:	1'014'174 716'159	TRWnom:	98'570

Table 70: Listing of all sign counts belonging to described flow item-sets in August 2007

8.4.6. Conclusions

The counted signs of the item-sets displayed in table 70 show, that the item-sets are describing a large amount of ICMP flows, which do not have the *GreyIP* sign set. The results do not help any further in classifying the observed flow item-sets.

8.4.7. Signs per Item-set February 2008

Period	February 2008			
Total Flow Count	3031759			

Item-set	**#118** IIP:192.33.90.66 IPort:0 prot:ICMP			
Flows matching item-set	359'365			
Counted Signs	ICMP: Onepkt:	359'365 231'541	Large:	9'104

Item-set	dOctets:244 dPkts:4 rPort:8 tos:0 IPort:0 prot:ICMP			
Flows matching item-set	366'520			
Counted Signs	GreyIP: ICMP:	89'998 366'520	TRWnom:	1'040

Item-set	dOctets:56 dPkts:1 rPort:11 durMs:0 IPort:0 prot:ICMP			
Flows matching item-set	422'621			
Counted Signs	GreyIP: Onepkt:	14'366 422'621	ICMP:	422'621

Item-set	dOctets:122 dPkts:2 rPort:8 tos:0 IPort:0 prot:ICMP			
Flows matching item-set	539'748			
Counted Signs	GreyIP: ICMP:	387'817 539'748	TRWnom:	110'903

Item-set	rPort:11 dPkts:2 IPort:0 prot:ICMP			
Flows matching item-set	303'415			
Counted Signs	GreyIP:	7'337	ICMP:	303'415

Item-set	rPort:11 tos:0 IPort:0 prot:ICMP			
Flows matching item-set	699'741			
Counted Signs	GreyIP: Onepkt:	14'909 281'441	ICMP: Large:	699'741 45'512

Table 71: Listing of all sign counts belonging to described flow item-sets in Febraury 2008

8.4.8. Conclusion

Item-set {**#118**} is the only one, having an IP address in it. This IP address belongs to the Planet Flow network and therefore has a benign cause. This item-set does not have the *GreyIP* sign and the Backsc sign set. This hardens the conclusion of a benign cause.

As expected due to the results of the sign occurence over the whole interval, table 71 does only show small amounts of the *GreyIP* sign and no counts of the Backsc sign in all item-sets.

8.5. Class Benign P2P

8.5.1. Over all Sign Statistics

Period	August 2007			
Total Flow Count	1'245'372			
Counted Signs	P2P:	1'245'372	PotOk:	216'102
	TCP:	253'234	TRWnom:	235'686
	Onepkt:	502'502	Artef:	2'816
	UDP:	992'138	Backsc:	1'930
	Large:	37'753	Retry:	642

Table 72: Count statistics over all signs occuring in interval of period August 2007

Period	February 2008			
Total Flow Count	4'192'548			
Counted Signs	P2P:	4'192'548	Artef:	5'055
	UDP:	4'015'363	PotOk:	464'063
	TCP:	177'185	Large:	863'831
	TRWnom:	109'683	Backsc:	2'782
	Onepkt:	1'597'532	Retry:	1'274

Table 73: Count statistics over all signs occuring in interval of period February 2008

8.5.2. Conclusions of Interval Sign Statistics

- what about trwnom

Tables 72 and 73 show the statistics over the inspected intervals. The listings show small amounts of the *Backsc* sign in each period. They are not assigned to the class “Backscatter” because they have an additional *P2P* sign. The *Backsc* sign is set, when only one packet is exchanged to a host on a well known port within 30 minutes. Flows having these signs set, could indicate traffic for updating a peer list or to check if the application still responds. The large amount of the *Onepkt* sign could also indicate flows which check if the application is still alive. When looking at the layer 4 protocols, it can be observed, that the UDP protocol is more used by P2P applications in the inspected intervals. In August 2007, UDP has a occurrence of 79,7 percent, in February 2008 it has even a occurrence of 95,8 %. Observing the *PotOk* sign shows, that 17,34 % in August 2007 and 11,1 % in February 2008 of the flows assigned to this class have IP pairings which have otherwise bidirectional communications.

8.5.3. Signs per Item-set August 2007

Period	August 2007
Total Flow Count	1'245'371

Item-set	rPort:31415 flowtype:10 tos:0 prot:UDP			
Flows matching item-set	141'539			
Counted Signs	TRWnom:	141'383	PotOk:	141'539
	P2P:	141'539	UDP:	141'539
	Onepkt:	85'724	Large:	3'648

Table 74: Sign counts per item-set in interval of period August 2007

8.5.4. Conclusions

The item-set described in table 74 shows, that almost all of the flows have the *TRWnom* sign set. All flows of the inspected item-set have the sign *PotOk* set, which leads to the conclusion, that these flows describe benign traffic. The observation of *Onepkt* signs is normal in P2P communication.

8.5.5. Signs per Item-set February 2008

Period	February 2008			
Total Flow Count	419'2547			
Item-set	**#119** dOctets:46 IIP:82.130.102.218 tos:0 prot:UDP dPkts:1 durMs:0			
Flows matching item-set	429'474			
Counted Signs	P2P:	429'474	PotOk:	126'034
	Onepkt:	429'474	TRWnom:	547
	UDP:	429'474	Backsc:	114
Item-set	**#119** dOctets:46 durMs:0 IIP:82.130.102.161 tos:0 prot:UDP dPkts:1			
Flows matching item-set	530'134			
Counted Signs	P2P:	530'134	PotOk:	159'566
	Onepkt:	530'134	TRWnom:	632
	UDP:	530'134	Backsc:	123
Item-set	**#119** dOctets:46 durMs:0 IPort:4246 tos:0 dPkts:1 prot:UDP			
Flows matching item-set	630'781			
Counted Signs	P2P:	630'781	Backsc:	237
	Onepkt:	630'781	TRWnom:	407
	UDP:	630'781		
Item-set	IIP:82.130.102.218 IPort:4246 tos:0 prot:UDP			
Flows matching item-set	1'049'880			
Counted Signs	P2P:	1'049'880	Retry:	250
	Onepkt:	304'980	TRWnom:	893
	UDP:	1'049'880	Backsc:	329
	Large:	309'212		
Item-set	**#119** dPkts:1 durMs:0 IIP:82.130.102.161 IPort:4246 prot:UDP			
Flows matching item-set	420'376			
Counted Signs	P2P:	420'376	TRWnom:	295
	Onepkt:	420'376	Backsc:	132
	UDP:	420'376		
Item-set	IIP:82.130.102.161 IPort:4246 tos:0 flowtype:2 prot:UDP ASlocal:559			
Flows matching item-set	1'315'844			
Counted Signs	P2P:	1'315'844	TRWnom:	1'027
	Onepkt:	370'766	Retry:	299
	UDP:	1'315'844	Backsc:	367
	Large:	391'311		

Table 75: Count statistics over all signs occuring in interval of period February 2008

8.5.6. Conclusions

The item-sets listed in table 75 show some *Onepkt* signs, which is absolutly normal in P2P traffic. The other characteristics in the counted signs are very small, therefore they have a lesser reputation. The observed IP addresses 82.130.102.161 and 82.130.102.218 belong to fake e-mule servers, which were active in this period. This could be the indication, why only a small amount of the *PotOk* sign is present per item-set. Each flow of the item-sets {**#119**} consists out of one packet, which indicates failing connection attempts or the anouncement of new data to the server.

8.6. Class Unreachable

8.6.1. Over all Sign Statistics

Period	August 2008			
Total Flow Count	4'138'700			
Counted Signs	P2P:	6'809	TRWnom:	237'913
	Unreach:	4'138'700	UDP:	4'076'891
	TCP:	61'809	Large:	89'811
	Onepkt:	1'004'674	Retry:	555
	PotOk:	91'078		

Table 76: Sign counts over the whole inspected interval in August 2008

Period	February 2009			
Total Flow Count	6'217'340			
Counted Signs	P2P:	35'210	TCP:	139'329
	Unreach:	6'217'340	Large:	191'626
	Onepkt:	706'076	TRWnom:	273'379
	PotOk:	246'943	Retry:	100
	UDP:	6'078'011		

Table 77: Count statistics over all signs occuring in interval of period February 2009

8.6.2. Conclusions of Sign Statistics over whole intervals

The statistics displayed in tables 76 and 77 show, that only a small amount of the traffic classified as unreachable uses the TCP protocol (1,5 % in August 2008 and 2,2 % in February 2009). Small amounts of potentially benign flows can be observed in both intervals. Flows having this sign set are failing connection attempts. Also shown in the statistics is a small amount of P2P traffic, targeting a host, which provides benign services, in a peer-to-peer behavior. A bit curios is, that in August 2008 24,3 % and in February 2009 11,4 % of the observed flow have a *Onepkt* sign set. All other flows contain more than one packet but can not be retries because the *Retry* sign is only showing up in a small amount. This could indicate a application trying to send a larger amount of data and not checking, if the receiving peer is alive.

8.6.3. Signs per Item-set August 2008

Period	August 2008			
Total Flow Count	4'138'700			

Item-set	dOctets:304 dPkts:4 ASremote:5432 rPort:123 IIP:129.132.2.21 lPort:123 prot:UDP tos:0			
Flows matching item-set	425'911			
Counted Signs	Unreach: UDP:	425'911 425'911	TRWnom:	2

Item-set	**#120** dOctets:76 ASremote:5432 rPort:123 tos:0 dPkts:1 durMs:0 IIP:129.132.2.21 lPort:123 prot:UDP			
Flows matching item-set	467'808			
Counted Signs	Unreach: Onepkt:	467'808 467'808	UDP:	467'808

Item-set	dOctets:380 dPkts:5 ASremote:5432 rPort:123 tos:0 IIP:129.132.2.21 lPort:123 prot:UDP			
Flows matching item-set	1'652'600			
Counted Signs	Unreach:	1'652'600	UDP:	1'652'600

Table 78: Sign counts per item-set in inspected interval of period August 2008

8.6.4. Signs per Item-set February 2009

Period	February 2009			
Total Flow Count	6'217'340			

Item-set	**#121** dOctets:152 dPkts:2 IIP:129.132.2.21 tos:0 lPort:123 prot:UDP			
Flows matching item-set	683'777			
Counted Signs	Unreach: UDP:	683'777 683'777	TRWnom: Retry:	707 65

Item-set	dOctets:760 dPkts:10 ASremote:5432 rPort:123 tos:0 IIP:129.132.2.21 lPort:123 prot:UDP			
Flows matching item-set	1'557'610			
Counted Signs	Unreach: UDP:	1'557'610 1'557'610	TRWnom:	4

Table 79: Sign counts per item-set in inspected interval of period February 2009

8.6.5. Conclusions over both periods

All of the observed item-sets listed in the statistics of tables 78 and 79 are using the UDP protocol. One item-set {**#120**} shows that all flows only consist out of one packet. Therefore this item-set represents failing connection attempts without any retries. The signs of all other item-sets indicate also failing connection attempts. Item-set {**#121**} has a small amount of flows having the Retry sign set, indicating more than one connection attempt per IP pairing.

9. Results and further Work

9.1. Results

The FIM analysis over a randomly chosen interval of three continuous hours chosen from a period, showing a significant peak in assigned flows, has shown, that the IBR Detector is classifying the inspected flows right. No item-sets were found, which should have been assigned to another class. This result does not mean, that all flows assigned to a class are assigned to the right class.

To verify, if the item-sets found in the analysis of a random three hour interval, a analysis over the whole peak periods is made. The results of this analysis show, that the randomly chosen interval was generating appropriate results in some of the inspected classes. This analysis has also brought up the cause of some of the peak periods. The table 80 lists all classes whose peak couldn't be identified.

Class	**Period**
Backscatter	February 2007
Benign P2P	August 2007
Benign P2P	February 2008
Unreachable	February 2009

Table 80: Listing of all classes whose peaks could not be determined.

Not observing the peak could have various causes. One of them is, that the peak could be caused by a lot of item-sets having a support smaller than 5%. According to a statistic of assigned flows to a class, the peak in period February 2007 of the class backscatter lies out of the inspected range. The data was not inspected, because of a time shortage.

The analysis of the reference intervals of periods without a peak have shown, that new flows, having other signs set, are not always the cause of the peak. Therefore, some reference intervals show the same item-sets as observed in the analysis of the intervals out of a peak period.

The sign statistics over the continuous three hour intervals with a support of 10 % has revealed the effectiveness of rules in the classes Backscatter and Other Malicious. It can be observed, that the flows in the class Backscatter are mainly assigned through the rule only having the *ICMP* sign in it. The rule having the *Backsc* sing in it, is the least effective. It only matches lesser than 0.1 % of the flows to this class. The rule with the *GreyIP* sign is somewhere in the middle of the other two rules.

Out of the inspected interval, the effectiveness of the rules from the class Other Malicious are calculated. This statistic shows, that 88.8 % of the flows are assigned by the rule containing the *Onepkt* sign.

9.2. Further Work

A lot has been done, but there is still a lot of work which could not be done. First of all, a analysis over all periods would give a better picture over the behavior of the one-way flows, because the inspected reference intervals are very short and do not cover all flow and sign item-sets belonging to the analyzed period.

Another analysis which could be done is to compare a peak period of a class with the class “Malicious scanning” to inspect if some flows are classified in the class with peak instead of the class “Malicious scanning”. This analysis would reveal some changes in item-sets, if the class with peak “boroughs” flows of the class “Malicious scanning”.

To gain a better overview over the sign occurrences in general, a statistic over the whole periods could be made. This would assure, that the statistic made is representative. Performing this analysis would especially help in determining the effectiveness of rules from classes having more than one rule.

Furthermore a sign statistic over all occurring signs per found item-set of a class could be made. This analysis would help to inspect the cause of a flows matching a given item-set.

10. Appendix

10.1. Bibliography

[1]	Scripts of the module Computer Networks 1 from HSR, written by Prof. Dr. Peter Heinzmann, Prof. Dr. Andreas Steffen, Prof. Beat Stettler
[2]	Protocol Numbers assigned by IANA: http://www.iana.org/assignments/protocol-numbers/protocol-numbers.xml
[3]	IP Protocol List on wikipedia: http://en.wikipedia.org/wiki/List_of_IP_protocol_numbers
[4]	TCP header image source: http://en.wikipedia.org/wiki/Transmission_Control_Protocol
[5]	RFC 3168: http://tools.ietf.org/html/rfc3168
[6]	TCP connection establishment and tear down picture source: http://de.wikipedia.org/wiki/Transmission_Control_Protocol
[7]	Steven Northcutt et al, Inside Network Perimeter Security, 2. Edition, Sams Publishing, ISBN 0-672-32737-6
[8]	IANA ICMP types and codes: http://www.iana.org/assignments/icmp-parameters
[9]	Eduard Glatz: „Visualizing Host Traffic through Graphs“ Sept. 14, 2010. Available at ACM: http://portal.acm.org/citation.cfm?id=1850802
[10]	Image Source HAP viewer role summarization: http://www.vizsec2010.org/files/Eduard%20Glatz.pdf, Page 6.
[11]	Eduard Glatz, Cenofontas Dimitropoulos: Beyond Network Telescopes: New Directions based on One-Way Flow Classification, ETH Zurich.
[12]	Konstantinos Karampogias: Evaluating and Improving the Detection of Internet Background Radiation. Feb. 2011, semester thesis at ETH Zurich.
[13]	Cisco Netflow format: http://www.cisco.com/en/US/docs/ios/solutions_docs/netflow/nfwhite.html
[14]	Configuring Netflow Data export: http://www.cisco.com/en/US/docs/routers/7600/ios/12.2SXF/configuration/guide/nde.html Table 51-3, Footnote 4
[15]	J. Jung, V. Paxson, A. Berger and H. Balakrishnan, Fast portscan detection using sequential hypothesis testing, in Proceedings of the IEEE Symposium on Security and Privacy, 2004, pp. 211-255
[16]	M. Allman, V. Paxson and J. Terrell, A brief history of scanning, in Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, 2007, p. 82
[17]	Sam source code: http://www.borgelt.net/sam.html
[18]	The software written by E. Glatz can be found in the folder software contained in the root directory of this project.
[19]	Info over AS4143 http://bgp.he.net/AS4143
[20]	Sober information from avira: http://www.avira.com/en/support-threats-description/tid/361/worm_sober.e.html/
[21]	Sans on port 37: http://isc.sans.edu/port.html?port=37

[22]	Newsgroup on ICMP code 8: http://www.mail-archive.com/nfsen-discuss@lists.sourceforge.net/msg01371.html
[23]	Planet flow labs: http://planetflow.planet-lab.org/
[24]	RFC 791: http://tools.ietf.org/html/rfc791
[25]	http://www.secudb.de/~seuffert/xbslink/faq#question-21
[26]	IP space of Swiss Federal Institute of Technology in Lausanne: http://www.db.ripe.net/whois?searchtext=128.178.176.86&searchSubmit=search
[27]	MMORPG port infos: http://forum.portforward.com/YaBB.cgi?num=1257588778/1
[28]	MMORPG UPnP infos: http://heroescommunity.com/viewthread.php3?TID=12141&pagenumber=1
[29]	IP 134.21.2.59: http://www.db.ripe.net/whois?form_type=simple&full_query_string=&searchtext=134.21.2.59&do_search=Search
[30]	Marcell Perényi, Trang Dinh Dang, Andras Gefferth, Sandor Molnar: Identification and Analysis of Peer-to-Peer Traffic, December 2006
[31]	http://www.db.ripe.net/whois?form_type=simple&full_query_string=&searchtext=as3215&do_search=Search
[32]	Service assignment to port number 9001: http://www.speedguide.net/port.php?port=9001
[33]	ETL manager: http://en.wikipedia.org/wiki/Extract,_transform,_load
[34]	Tor informations: https://trac.torproject.org/projects/tor/wiki/TheOnionRouter/TorFAQ#SoIcanjustconfigureanicknameandORPortandjointhenetwork
[35]	It is not possible to link to a search result, because they do not have unique URLs. Therefore only homepage: www.apnic.net
[36]	Regional Texan ISP on ARIN: http://whois.arin.net/rest/net/NET-67-18-0-0-1/pft]
[37]	Distributed Computer Group @ETH Zurich: http://www.disco.ethz.ch/
[38]	Uplink network: http://www.uplink-netz.ch/index2.html
[39]	MS Knowledge Base article on SQL server ports: http://support.microsoft.com/kb/287932
[40]	RFC 863 at IETF: http://tools.ietf.org/html/rfc863
[41]	Optika eMedia Service: http://en.wikipedia.org/wiki/Oracle_Imaging_and_Process_Management
[42]	Info over services running on port 11000: http://www.speedguide.net/port.php?port=11000
[43]	Senna Spy Trojan generator: http://www.symantec.com/security_response/writeup.jsp?docid=2001-062211-2540-99&tabid=2

10.2. Glossary

Term	Description
(D)DoS	Distributed Denial of Service Attack. Within DDos attack a service or host is being flooded with an enormous amount of requests. The goal of such an attack is to temporarily make the service or host unavailable. The requests are sent from different clients spread over a network or the Internet.
Flow (data)	Describes a network connection. Flows can be uni- or bidirectional.
Packet Header	Packet Headers include the information for routing a packet through a given network.
IDS	Intrusion Detection System. This is a computer system which inspects network traffic and generates alarms if given patterns are found in the inspected traffic. The patterns are describing the behavior of possible attacks.
DPI	Deep Packet Inspection. The technique of inspecting packet data to identify malicious traffic.
Frequent Item set Mining	This method takes a given attribute of any item and records, which other attributes are set and how often they appear.
TCP Flags	TCP flags are located in the TCP header and are used to establish, tear down or control a TCP session.
Bogon IP	An IP address or address space which is at the moment not assigned by IANA or one of the regional registrars.
Backscatter	In this terms, backscatter is related to traffic, which uses a spoofed IP address. This IP could be free, as it is not assigned to a host, or out of the private IP ranges, which are not routed on the Internet.
Bogon	Bogon addresses are faked IP addresses, claiming to come from IP space not assigned yet or which is known to be not in use.
P2P	Applications which do communicate without the assistance of a server are called peer-to-peer applications.
Cluster	A cluster is a composite of some workstations or servers which are connected through a fast network infrastructure or have a common file storage.
DNS	The Domain Name Service is responsible for resolving requested URLs into IP addresses.
SMTP	The Simple Mail Transfer Protocol is used to upload e-mails to an e-mail server.
E-mule	E-mule is a common peer-to-peer application for sharing files over the Internet.
Bit torrent	Bit torrent is like e-mule, but uses its own and for larger files faster protocol.
AS (Autonomous System)	An Autonomous System announces routing path information to the subnets assigned to it.