The global View of Security Situation in the Internet

The constantly growing importance of the Internet for our information society makes it necessary to analyze and be acquainted with its Security Situation beyond the limits of the individual network operators.

Prof. Dr. Norbert Pohlmann

Professor in the Computer Science Department for distributed systems and information security and director of the Institute for Internet Security at the University of Applied Sciences Gelsenkirchen, Germany.
Chairman of the board of the TeleTrusT association.
Member of the Permanent Stakeholders' Group of ENISA.
Chairman of the ISSE program committee.
www.internet-sicherheit.de
www.if-is.net

We have all experienced the situation: you are sitting in a traffic jam and all you can see is a long line of cars in front of and behind you. In this situation, without any assistance, you do not have an overview of the problem. There is no direct information concerning why the traffic jam has come about, how long it is, at what point of the traffic jam you are located or - the most important information - when the traffic jam will dissolve. As this is a problem faced on a day-to-day basis by thousands of motorists, solutions have been developed to overcome the lack of information. There is a close network of traffic counter loops which record the traffic volume and situation on the motor-ways/freeways. Important information about traffic jams is provided by means of radio announcements, SMS, telephone and the Internet, while modern navigation systems process the information directly when planning the route to be taken. Through the use of these resources, motorists are "liberated" from their constricted local view of the situation and can take decisions in good time on the basis of the global information available, e.g. leaving by the next exit and using an alternative route.

This situation can also be applied to the perspective that the network operators have of the Internet today. As a rule they have only a local perspective, i.e. an overview of their own network segments and the communication data that is transferred. If problems occur here and are detected, they can be rectified quickly and systematically. However, if it becomes apparent that a problem has occurred that is not within their own domain of action, or if the required perspective is lacking, the situation is more difficult. In most cases we do not know the origin of the problem and we are reliant on third parties to solve the problem.

The global view for the right decision.

The global view of security situation in the Internet required in order to detect the problem and to select the appropriate solutions is missing. Such a global view on the Internet is difficult to achieve as people like to play their cards close to their chest. The precise internal network structure, communication connections and topologies are often treated confidentially by the network operators [1].

Furthermore, in order to obtain a global perspective, there are a few challenges that have to be coped: communication data is relevant in principle to data protection, the quantities of data are enormous, the data rates are sometimes so large that they cannot always be analyzed in real time, while long-term storage of the communication data in order to observe long-term developments appears to be impossible. Moreover, the question also arises of who feels responsible for creating a global perspective?

Nevertheless, the Internet has developed into an omnipresent medium over the

past few years, without which very large areas of the economy, research and private life would be unimaginable today. For this reason the analysis and knowledge of the medium known as the Internet in its totality is of particular significance in order to be able to assess its development and guarantee the future functioning of all the services it provides.

The constantly growing importance of the Internet for our knowledge and information society makes it necessary to analyze and be acquainted with its status beyond the limits of the individual network operators. Only precise knowledge of the normal status makes it possible to detect anomalies which influence the functionality of the Internet.

With the help of the probe-based Internet Analysis System, which is currently being implemented as a research and development project of the Institute for Internet Security at the University of Applied Sciences in Gelsenkirchen in collaboration with the German Federal Office for Information Security (BSI), it is intended to create and analyze local and above all global perspectives in order to make the generation of the global view of the security situation in the Internet possible.

Aims and Task of the Internet Analysis System

The task of the Internet Analysis System on the one hand is to analyze local communication data in defined subnetworks of the Internet, and on the other to create a global perspective of the Internet by bringing together the large number of local perspectives.

The functions of the Internet Analysis System can be divided up into the four subsegments of pattern formation, description of the actual status, alarm signaling and forecasting.

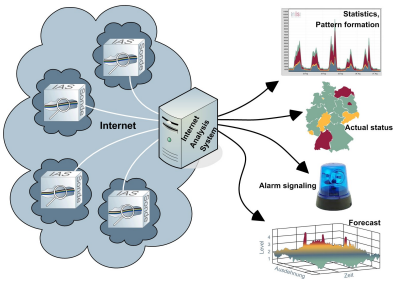

Fig. 1: Tasks of the Internet Analysis System

The main task of pattern formation is a comprehensive analysis and interpretation of the communication parameters of Internet traffic, with the aim of detecting technology trends, interrelationships and patterns which represent the various statuses and perspectives of the Internet. On the basis of this knowledge a search is carried out for anomalies among the current measured values and the causes of status changes analyzed and interpreted. Here it is important to find out whether the status anomalies have a natural origin, for example as a result of a technological change, or whether they are attributable to a wanton attack.

With knowledge of the current status of a communication line and the use of historical - i.e. previously collected - information (knowledge base) it is possible in the case of significant changes to traffic volumes or communication data to generate a warning message, on the basis of which measures can be initiated to protect and maintain the correct functioning of the Internet.

If you can‘t measure it, you can‘t manage it!

A further important function is the visual depiction of the Internet status similar to a weather or traffic jam map. Here intuitive depictions are being developed with which the most important parameters are discernible at first glance.

Fig. 2: Security Situation in the Internet

Through the examination and analysis of the extrapolated profiles, technology trends, interrelationships and patterns it will be possible by means of an evolutionary process of the acquired results to make forecasts of Internet status changes. In this manner it is possible to detect indications of attacks and important changes at an early stage and forecast the effects of the damage.

Principle of Raw Data Collection

Figure 3 shows the principle of raw data collection by the probes. This is divided up into three sections. The Internet is represented on the left. Packets of three different application sessions are shown: related HTTP packets, an FTP session and an SMTP session. The probe is located in the middle of illustration 3. The packets of the three applications are accessed passively by the probe one after the other in their random order and evaluated. The packet that is accessed is channeled through several analysis categories, each of which is responsible for a certain protocol. These evaluate strictly defined communication parameters in the protocol header at the various communication levels which are not relevant to data protection law. The counters allocated in the counting system are incremented according to how the header information of the

packet is filled out. The frequency of certain header information is recorded in the same way as on a tally sheet. For simple example, in illustration 3 the accessing of the FTP packet is recorded by incrementing the FTP counter by 1. The raw data are therefore aggregates of counters, i.e. counters of communication parameters that have appeared at the various communication levels over a defined period. The packet - in illustration 3 an FTP packet - is immediately deleted physically, i.e. irreversibly and without trace, by the probe [2].

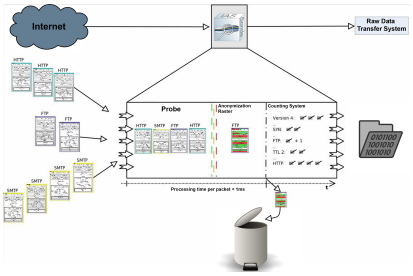

Fig. 3: Principle of raw data collection

Reconstitution of the context of a packet or only a communication parameter is not possible or necessary. At definable intervals the counter readings (raw data) of the probes can be transmitted to the raw data transfer system. All of this information is completely anonymous, as shown in Figure 4.

ID	Description	Count
131134	IP (Protocol Number 6)	: 18.854.151
131145	IP (Protocol Number 17)	: 1.123.149
327708	TCP (Flags: SYN)	: 334.435
327723	TCP (Flags: FIN/ACK)	: 480.697
327724	TCP (Flags: SYN/ACK)	: 275.779
545857	HTTP (Request Method POST)	: 2.026
545861	HTTP (Request Method GET)	: 293.616
545863	HTTP (Request Method HEAD)	: 18.992

Fig. 4: Counting system in the probe

On the right after the colon are the

counter readings for the header information specified on the left. Each line stands for a counter. On the left-hand side of the colon is the count-if function (appearance of the corresponding communication parameters) and on the right the number of packets which contained the communication parameter during the defined measurement period. For example, line 2 of the raw data shown indicates that 1,123,149 packets with the IP protocol number 17 (UDP) appeared in the prescribed time.

Some results of the Internet Analysis Systems

For the purposes of illustration some results are presented in this section in order to provide an idea of the possibilities of the current status of the Internet Analysis System. At present there are approximately 800,000 different counters of communication parameters incorporated for the various communication levels. This large number clearly shows how complex the results can be.

Types of E-mail Messages

Illustration 5 shows the ability of the system to record the statistics of the headers of the e-mails sent by SMTP. The distribution can provide information on general communication behavior, as well as deviations from it.

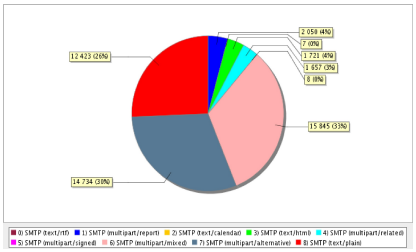

Fig. 5: Distribution of e-mail Content Types

Figure 5 shows an example of normal behavior in which the total number of messages without attachments represents 60% of all messages. These e-mails include messages with the text/plain, text/html and multipart/alternative content types. As a rule, e-mails with attachments are provided with the multipart/mixed content type. A mixed form is e-mails with the multipart/related content type. Here, for example, images are integrated directly into the text. If these e-mails are included in the e-mails with an attachment, approximately 36% of all e-mails are sent with an attachment. The remaining 4% essentially consist of confirmations of reading with the multi-part/report content type. An abrupt change to these values in particular may indicate a wave of spam affecting a company from the outside, or indicate that a computer is sending spam from within the company.

All of this information is anonymous.

Transport Protocol Distribution

Figure 6 shows the distribution of the protocols of the transport layer used over a period of several days for a specific communication line.

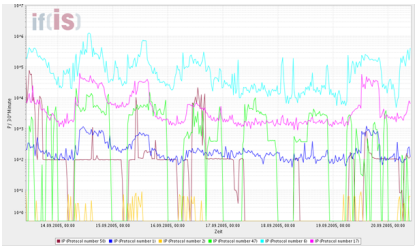

Fig. 6: Protocols of the transport layer

From the past the Internet Analysis System knows the profile, the standard deviation and from this can display an indication of untypical behavior. Additionally, the use of certain protocols can be determined, enabling capacity planning for the use of Virtual Private Networks (ESP protocol), for example. Protocol dependencies can also be detected: UDP appears to be proportional to TCP, which can be attributed to the dependencies of HTTP and DNS.

More aspects

In the area of research various universities are working on other important issues, like the recognition and analysis of Trojan horses, pattern

recognition, detection of anomalies, neural network models for communication parameters, DataMining algorithms, and anonymization.

It has to be analyzed on which spots of the internet the probes need to be placed to make a representative statement [3].

On the sensors level there are more systems, like log-data based systems that have access to router log data, switches, Intrusion-Detection-Systems, firewalls, web servers and therefore are able to analyze it.

Examples for such systems are the "Symantec DeepSight Threat Management System" and "DShield.org - Distributed Intrusion Detection System".

The driving force behind DShield is the Internet Storm Center of the SANS Institute in the USA. Anybody who operates a firewall and is willing to contribute his logfiles to the project can participate. The possibility to contribute logfiles completely anonymous, without checking the country of origin and time zone is quite questionable. This raises a considerable doubt on the trustworthiness of data.

Most of the input data comes from the USA. Together with designated experts important alerts can be spoken out by the system.

Other Early-Warning-Systems have an active access to internet services and record the availability data. This enables a quick overview of the availability of important services like DNS, E-Mail, and web servers.

Besides the sensors and interpretation level there are a valuation and categorization level, which can refer recognized irregularities to normal network behaviour or a cyber attack.

The valuation and categorization can hardly be automated.

This is the point where humans have to take decisions, based on their technical knowledge, their expertise and by accessing additional information.

Another very important aspect is the distribution level. Here the addressees of alerts have to be selected carefully. If responsibilites and competence fields are't defined exactly, the alert may be sent to an addressee, who is not allowed to react, or doesn't have the suitable knowledge to perform the necessary measures.

Perspective

Even if we do not know today if we could recognize the most important attacks, we need to be able to have a global view of the internet.

Similar to the situation of road traffic the results will be implemented to infrastructural security measures (Black List, Router Polies, Identity Management, and so on) and to a higher level of operating system security (Trusted Computing, and so on.), as well as applications (e.g. digitale signature).

We have to make ourselves aware that there are new laws to come, which will help to create a more trustworthy internet.

We face a challenging way to establish a working Internet-Early-Warning system.

We face a challenging way to establish a working Internet Early Warning System.

The cooperation of companies, organizations and governments is important to create a global view of the internet.

By that we will be able to detect attacks in time.

Further information

Institute for Internet Security, http://www.internet-sicherheit.de or http://www.if-is.net

Federal Office for Information Security (BSI), http://www.bsi.de/english/index.htm

References

[1] N. Pohlmann, M. Proest: „Internet Early Warning System: The Global View", in "Securing Electronic Business Processes - Highlights of the Information Security Solutions Europe 2006 Conference", Publisher: S. Paulus, N. Pohlmann, H. Reimer, Vieweg-Verlag, Wiesbaden 2006

[2] N. Pohlmann: "Probe-based Internet Early Warning System", ENISA Quarterly Vol. 3, No. 1, Jan-Mar 2007

[3] S. Dierichs, N. Pohlmann: "Netz-Deutschland", iX - Magazin für professionelle Informationstechnik, Heise-Verlag, 12/2005

[4] N. Pohlmann: "Internetstatistik", Proceedings of CIP Europe 2005, Publisher: B.M. Hämmerli,