Internet Situation Awareness

Malte Hesse
Institute for Internet Security
University of Applied Sciences Gelsenkirchen
Neidenburger Str. 43
45877 Gelsenkirchen, Germany
+49 209 9596 763
hesse@internet-sicherheit.de

Norbert Pohlmann
Institute for Internet Security
University of Applied Sciences Gelsenkirchen
Neidenburger Str. 43
45877 Gelsenkirchen, Germany
+49 209 9596 515
pohlmann@internet-sicherheit.de

ABSTRACT
The Internet is consisting of autonomous systems each managed by individual and rival organizations, which makes it very difficult to capture as a whole. An Internet Situation Awareness can be accomplished by creating a common basis for private and public operators to monitor their networks, by offering them a common smart approach and the additional benefit of a global view, which they can use to compare their local situation with. This smart approach should utilize well proven existing global statistics, best practices and existing technical sensors, which can be adapted to the overall common framework. From this, output for all relevant stakeholders, like national assessments centers, can be generated to fulfill the individual needs.

One possible input source could be the technical sensor technology, which has been developed by our Institute and which we give to partners and other researchers free of charge. It is a great basis for an Internet Situation Awareness, since it is a well proven system, which has been in operation for a couple of years, and since it can easily be adapted by our developers to comply with the overall framework. The great advantages are in addition (i) that it is privacy compliant by design and (ii) can offer high performance with the (iii) capability for long time storage of the collected raw data. Using raw data collected at various positions of the internet infrastructure, we aim to generate a continuous global view of the current state of the Internet, which can be utilized as input for the Internet Situation Awareness.

Categories and Subject Descriptors
C2.0 [**Computer-communication networks**]: General – *data communications, security and protection.* C2.3 [**Computer-communication networks**]: network operations – *network management, network monitoring, public networks.* K6.m [**Management of computing and information systems**]: Miscellaneous – *Security.*

General Terms
Management, Measurement, Economics, Reliability, Security, Legal Aspects.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Conference'04, Month 1–2, 2004, City, State, Country.
Copyright 2004 ACM 1-58113-000-0/00/0004…$5.00.

Keywords
sensor network, internet situation awareness, global view, national assessment center, homeland security system, critical infrastructure protection, internet analyzing system, internet research, internet early warning

1. INTRODUCTION
The Internet has become a large and complex system, which goes beyond all geographical, political, administrative and cultural borders, leaving a new and unusual challenge to our society. Our society is currently undergoing fundamental changes cased by ever increasing connectivity and the penetration of the Information Society, in which the Internet with its plethora of services plays an important role. By now some parts of the Internet have secretly emerged to be critical assets. The importance is yet still growing, due to the convergence of fixed and mobile networks and evolving all-IP concepts. The Internet is not regulated and consists of self-governed Autonomous Systems (AS) each managed by individual organizations mostly part of the private sector. Currently there are about 27.000 different ASs advertised in the global routing table [3]. Private organizations are exposed to a high level of competition making an exchange of important management information between them impossible. The precise internal network structure, communication connections and topologies are often treated as being confidential by the network operators [20].

Besides the management of their own AS, the organizations also have to develop a strategy to exchange data with other AS. There are more than 60.000 logical connections between ASs at the moment [3]. Main factors for this routing of data are (i) the type of AS itself, like AS of large companies, Internet Service Providers, Universities and Internet Exchange Points, (ii) (company-)policies and (iii) - more evident - financial aspects. At the moment no one can say how much inter-European traffic is routed unnecessarily through non-European networks like through Russia just for minor financial reasons, leaving security aspects out of perspective. In the future it might be interesting, due to financial reasons, to transfer inter-US traffic by the means of Chinese providers. Thus, economical necessities affect the organization's proceedings, which yields to a reduction of redundancy and therefore to a destabilization of the internet infrastructure.

1.1 Lack of information
Currently there is a national discussion about the percentage of data traffic created by illegal downloads in comparison to the total traffic in Germany. Similar discussions are taking place all around

the world. This shows that the underlying structure of the Internet is very complex and that nobody can offer a global view presenting adequate details. At the moment every organization operating as part of the Internet simply has its own local view. Evaluation and analysis depends purely on the experience, resources and the used monitoring devices within each organization. Internet exchange points for instance can only offer information on traffic that is exchanged between AS. But how does the client program of a peer-to-peer network select its peers? Is that done mainly due to performance reasons and would these peers most probably be found within the same AS if available there or is it more likely to find the demanded content in Russia on servers with great connectivity?

ISPs earn money by transferring data, so they are not keen on providing evidence on how much money they make by routing illegal content. In addition, they can only offer their local view, which might differ from ISP to ISP, depending on their total number of end customers in relation to business customers. An exchange of management information between private partners is almost impossible, due to the high level of competition.

1.2 Internet Situation Awareness

The term Situation Awareness (SA) comes from the area of air traffic control and military command & control. The term is generally used when the understanding of an environment is critical for the process of decision making. A situation awareness of the Internet is - as described - very difficult to capture and until now nobody can offer a truly global view or even as much as a broad overview. This is leaving decision makers, regulators, users and service providers without a reliable foundation to base their decisions on.

In Endsly's model of situation awareness he described in 1995 three sequent phases performed in the process of situation assessment: (i) perception of elements in current situation, (ii) comprehension of current situation and (iii) projection of future status. These phases map to our efforts to generate a continuous Internet Situation Awareness, which will be described in this paper. To generate this SA input from all relevant but well selected and trustworthy sources will be used such as from statistics, technical sensors and partners as part of the situation assessment. The technical sensors must comply with an overall framework. Therefore, in most cases open source technology should be preferred.

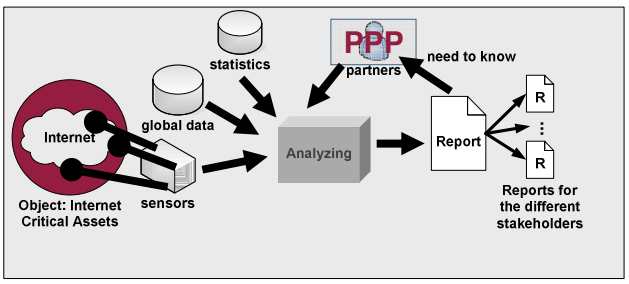

Fig. 1: project idea for a continuous Internet Situation Awareness

We also have to acknowledge the fact that monitoring Critical Infrastructures (CI) is a national competence in most regions of the world. Our goal is therefore to help establish a needed additional knowledgebase for the national governments for their assessments centers, to enable them to suggest and/or perform preventive actions and countermeasures.

Besides the output for the national governments additional output can be generated for private partners and citizens. Therefore, reporting has to meet various needs and perspectives. The output, we want to offer to interested partners can vary, is depending on particular requests.

1.3 Obtaining a global perspective

In order to obtain a global perspective, there are a few challenges that have to be coped with: (i) communication data is relevant in principle to data protection, (ii) the quantities of data are enormous, (iii) the data rates are sometimes so large, that they cannot always be analyzed in real time, while (iv) 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. According to the European Network and Information Security Agency (ENISA) up to 30 percent of the global trade are "digitally dependent" [11]. 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.

2. Sensor technology

Our intention is to create and analyze local and above all global perspectives in order to make the generation of the global view of the Internet possible. Therefore, we have developed a passive sensor technology, that can continuously collect statistical raw data with sensors, which are placed at selected spots of the internet communication infrastructure. At the moment we have implemented sensors mainly in Germany to monitor the internal government network and the networks of Universities and Companies. In addition, we have also found partners in Austria and Brazil.

The raw data is captured from header information of the passing network traffic, by counting the occurrence of (currently) 870.000 different parameters. This processing is assuring, that all sensitive header information, like IP addresses and user data, is left out and therefore the data is not privacy-sensitive, avoiding ethical, privacy, and legal challenges, that other data-collecting systems are plagued with. Additionally, our processing is designed to have very high performance and it allows frequently transfers of the collected data to our centralized database using encryption. This

enables us to collect and store data securely over a long period of time. The general approach is different form Intrusion Detection Systems, which only collect information in case of a specific exception event, and different from other monitoring systems, that store highly confidential content or IP addresses, which are privacy sensitive. These mentioned systems are well proven in the local environment, but cannot cope with the global challenge.

2.1 Aims and Tasks of the Internet Analysis System

The task of the Internet Analysis System on the one hand is to analyze local communication data in defined sub networks of the Internet, and on the other to create a global perspective of the Internet by bringing together the large number of local views. The functions of the Internet Analysis System can be divided up into the four segments of (i) pattern formation and creation of a knowledge base, (ii) description of the actual status, (iii) alarm signaling and (iv) forecasting.

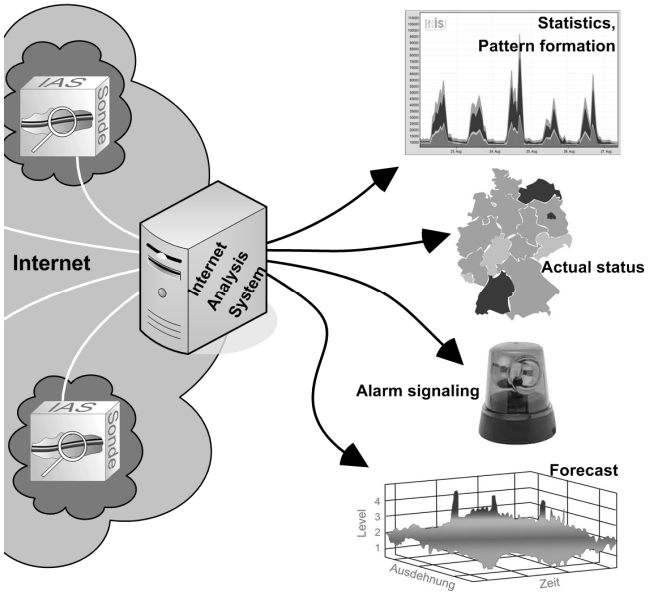

Fig. 2: Tasks of the Internet Analysis System

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. We also want to create a knowledge base, which we can use to understand the functioning of the internet from the "communication behavior" point of view.

On the basis of this knowledge a search is carried out for anomalies among the current measured values and the causes of status changes are 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 infrastructure 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. By this, we can limit the damage caused by a possible attack.

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 [26] (please take a look at the example in Fig. 3).

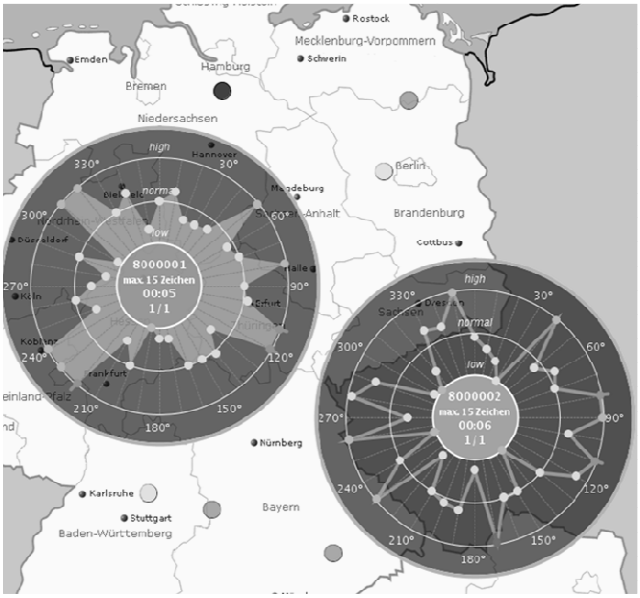

Fig. 3: Security Situation of the German Internet

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 changes of the Internet status. 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, which helps us to avoid this damage created by a possible attack..

2.2 Principle of Raw Data Collection

Fig. 4 shows the principle of raw data collection by the probes. This is divided up into three sections. (i) The Internet is represented on the left. In this example packets of three different application sessions are shown: related HTTP packets, an FTP session and an SMTP session. (ii) The probe is located in the middle of Fig. 4. 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 laws. (iii) 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.

Let us take a look at the following simple example: In Fig. 4 the accessing of the FTP packet is recorded by incrementing the FTP

counter by 1. The raw data is therefore an aggregation of counters, i.e. counters of communication parameters, that have appeared at the various communication levels over a defined period. The packet - in Fig. 4 a FTP packet - is immediately deleted physically, i.e. irreversibly and without trace, by the probe [21].

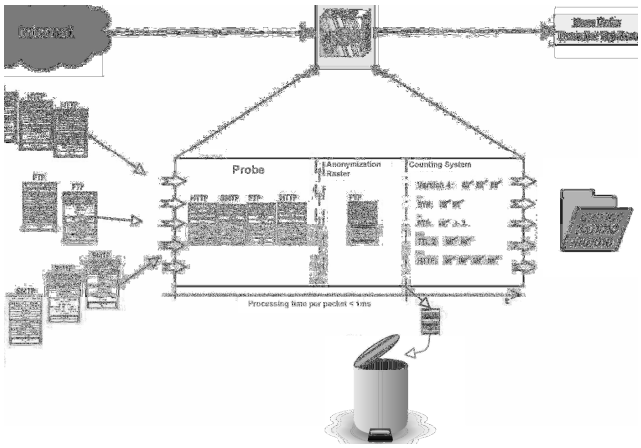

Fig. 4: Principle of raw data collection

The sensor monitors for instance the protocols: IP, ICMP, TCP, UDP, HTTP, SMTP, FTP DNS, EMULE, IRC, RTP, SIP, Skype, etc [25]. For each protocol a different number of monitored parameters is implemented in the sensor technology and therefore observed by it. The number is depending on the importance and distribution of the protocol. Therefore, we count for SMTP 1.624, for HTTP 1.123, for DNS more than 9.000, for RTP only 47 and for TCP 680.551 different parameters. The number for TCP is as high, because it includes the observation of the different ports as well. These counters sum up to a total of currently 876.596 different parameters, for which the occurrence is registered by the sensor [25].

2.3 Data privacy vs. data confidentiality

Speaking from our perspective, the European Union has very high standards for the privacy of the data of citizens. This was considered by the design of the sensor. Therefore, no portions of the communication packets, that could be linked to an individual, is further processed or stored. This includes the payload of the packets at the level of the application layer. Even IP addresses are considered to be off limits, due to the fact that the internet service provider can link the address to a customer (real person).

Reconstitution of the context of a packet or only a communication parameter is not possible or necessary. At intervals definable by the partner the counter readings (raw data) of the sensor can be transmitted securely to the raw data transfer system using encryption [2]. For most partners this interval is set to be five minutes. All of this information is completely anonymous, as shown in Fig. 5.

On the right behind the colon character are the counter values for each parameter specified on the left. Each line stands for one counter. For example, line 2 indicates that 1,123,149 packets with the IP protocol number 17 (UDP) appeared in the prescribed time interval. The total table has more than 870.000 lines.

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. 5: example of results of the counting system in the probe

One of the largest tradeoffs caused by complying with the privacy laws is the fact, that we cannot work with the IP address of a potential attacker and with the payload of the packets. Therefore, we can only locate the origin of the attack roughly by locating the sensor, that first recorded a phenomenon. In addition to this, we can estimate with a certain percentage, if it was a distributed or a centralized attack by using certain parameters.

And some attacks, that are hidden in the user data and aim to manipulate the server like the “Invite of Death (IoD)” for Private Branch Exchanges, cannot be detected [25]. Therefore, other systems, that are used by the operators of AS in case of the detection of a certain local event, offer an enrichment to the Internet Analysis System and are necessary to create a continuous Internet Situation Awareness.

Of course, one could argue scenarios, in which the Internet Analysis System could also be used in a privacy violating fashion. Recently we have passed out an enhanced DSL router to students on a voluntary basis. On top of the DSL router hardware we have implemented this described sensor in combination with another development of us, an active drone monitoring the availability of services out of the perspective of users [18]. Since we can link the raw data coming from each router directly to a student, this is of course a scenario, in which we need the consent of the concerned. So in all cases, in which the sensor raw data can be linked to one single individual, we have a data privacy problem. If it is not possible to get the consent of the concerned the sensor should only be placed in working environments with at least 50 people. But this DSL scenario is not what the sensor was designed for. The great advantages show when placed in the communication infrastructure of larger networks.

So we are proud to say, that we do not have to deal with privacy issues. But we are aware that by processing the raw data we might reveal information - like from economical nature -, that to some extent can prove to be critical. Therefore, we have to deal with data confidentiality issues. If one links the knowledge about a crisis in a product of a company, with the fact that there is no increased level of e-mail communication on the weekends, this might show, that the staff of the observed organization is not actively working on solving the problem. This is a similar situation to the DSL router scenario, but in this case it is not linked to an individual but to one single organization. We have a great understanding for the fact, that the confidentiality of the raw data of individual organizations is essential for the success of the whole idea. We encounter this by establishing trust and transparency on an organizational level and by using encryption and smart processing of data on a technical level. We need to make sure, that the data of an organization does not get in the hands of another directly or by the means of reverse engineering.

2.4 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 abilities of the current status of development of the Internet Analysis System. At present there are - as mentioned - approximately 870,000 different counters for communication parameters implemented on various layers of the communication. This large number clearly shows, how complex the results can be. Here, we now present some basic examples:

2.4.1 Types of E-mail Messages

Fig. 6 shows the ability of the system to record the statistics of the headers of the e-mails sent via SMTP. The distribution can provide information on general communication behavior, as well as deviations from it.

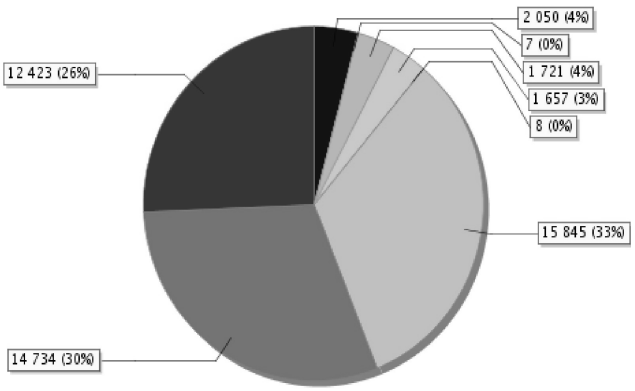

Fig. 6: Distribution of e-mail Content Types

Fig. 6 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 (12423), text/html (7) and multipart/alternative (14734) content types. By rule, e-mails with attachments are provided with the multipart/mixed (15845) content type. A mixed form is e-mails with the multipart/related (657) content type. Here, for example, images are integrated directly into the text. If these e-mails are included in the total count of e-mails, which are having 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 (2050) content type. An abrupt change of 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. It could also be an indication for an attack with malware attached to e-mails.

2.4.2 Transport Protocol Distribution

Fig. 7 shows the distribution of the protocols of the transport layer used over a period of several days for a specific communication line.

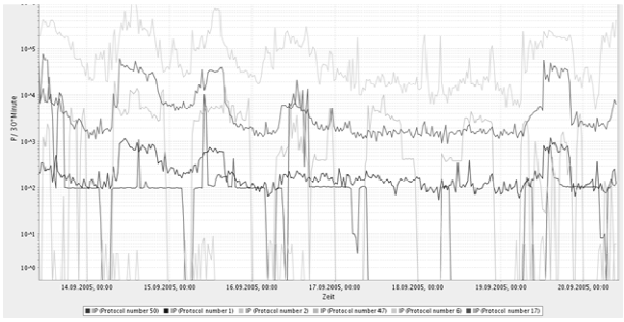

Fig. 7: Protocols of the transport layer

From past readings the Internet Analysis System has stored the profile of the standard deviation, which can be used to detect and 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 (using TCP) and DNS (using UDP).

2.4.3 TLS cipher suites really used for encryption

For the secure communication between clients and servers so called cipher suites have been pre-defined consisting of methods for the key exchange with authentication and algorithms for encryption as well as for data integrity. Which cipher suite is used for the communication is then negotiated between client and server depending on availability of algorithms and set preferences. If the browser is connecting the web server the browser offers the possible crypto suites to the server. Then the web server decides, which crypto suites should be used for the communication.

Since modern encryption is based on problems of complexity and due to the fact that weaknesses in some methods have been identified, some cipher suites should no longer be used with growing performance of the available computer systems. But sometimes the use of certain very insecure cipher suites is mandatory, due to questionable national laws, which are supposed to ensure legal interception especially in so called “rogue regimes”.

The Internet Analysis System can monitor from authentic network traffic, which cipher suites are really being used. This information is very interesting for all national representatives in charge for the monitoring of the infrastructure internet. So far they have to base their decisions on very little available information. Most of the time they are not aware of the actual situation.

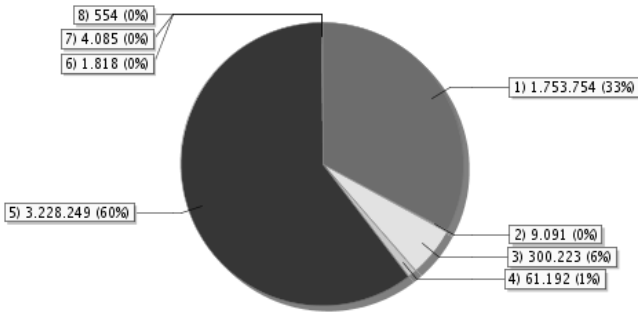

Fig. 8: Distribution of cipher suits used

In one of our monitored sub networks we have recorded the following distribution (Fig. 8): In 60% of all encrypted communication the very common RSA_WITH_RC4_128_MD5 (5) cipher suite was used, in 33% the improved and more secure DHE_RSA_AES_256_CBC_SHA (1) cipher suite and in about 6% the RSA_WITH_AES_128_CBC_SHA (3) cipher suite was used, which is from the perspective of security also fine. But we have detected some profiles, that should not be used, like in 0.1% of all encrypted communication in form of the RSA_EXPORT_WITH_RC4_40_MD5 cipher suite, which only offers a 40 bit key length or in the case of 0.01% of the encrypted communication in form of the RSA_WITH_NULL_SHA cipher suite, actually offering no encryption at all. From this information national representative can disseminate guidelines for a secure use of the internet for agencies, companies and citizens.

2.5 Further analyzing of the statistical raw data

The statistical raw data we are already collecting is almost certainly concealing very interesting information for a variety of scientific disciplines. Collecting raw data from a plethora of sensors will result to a huge amount of information. Analyzing our special type of statistical raw data will require new intelligent techniques and algorithms. A very interesting discipline will be to explore these techniques and in addition to investigate, which procedures from the areas of machine learning and artificial intelligence can be adopted for our purposes. We aim to search and discover important and critical information, that are not easily located by traditional methods. For example, by linking the absence or presence of parameters at certain events (using data mining for instance [29]). This is just a very basic example, unfortunately our resources at the moment only allow very little and focused research. Thus, we are very interested in extending our existing cooperation in this area with interested research institutes and we can offer a large amount of past and present data as well as the sensor and evaluation technology.

The focus is not simply on technological, security or performance related information, but we also want to provide a broader multidiscipline vision by including aspects form the disciplines like economy and sociology. The driving force behind this idea is the observation that similar environments exhibit different behaviors when using similar technologies. For example, we have observed that one of our sensors in Brazil placed at a university records different usage patterns than our sensors at German universities [25]. We believe this different usage of technology might have to do with sociological aspects and performing further research in this direction could provide stunning results. These results could for instance be very beneficial for companies looking for early adaptors to test a new technology.

2.6 Global View

As illustrated, the Internet Analysis System can be used to generate a local view of IP-based networks. These networks, ranging besides others from companies', Internet Service Providers', Content Providers', Universities' networks, come along with very different characteristics, like obviously the total number of packets passing the sensor. The local view already helps the operator to monitor the network, but a global view is even a lot more valuable [23][28]. It can be used to compare the local situation with an authentic global view to detect abnormalities, which might help to confirm or dismiss the detection of local attacks or events. The global view is valuable to a number of other relevant stakeholders as well, like for national assessment centers.

To generate this global view partners are invited to join by frequently sending a summary of their local view to a centralized evaluation system (Fig. 9). From this authentic data the common global view is generated and transferred back to the participating partners. Due to the structure of the internet this can only be accomplished with the support of the partners. So far nobody can offer this kind of global view, so you cannot just go out and buy it someplace.

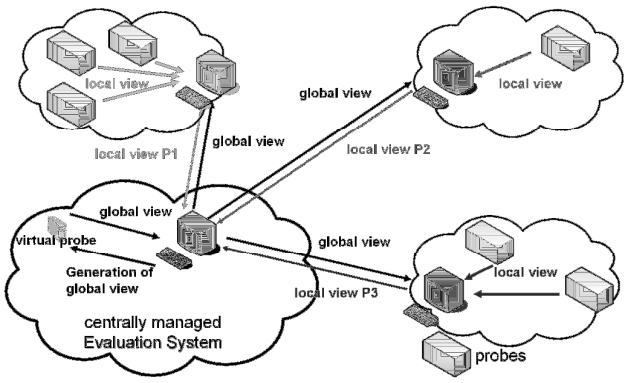

Fig. 9: collaboration for a global view

In Fig. 10 we give an example of the possible confirmation of an attack with malware attached to e-mails by the use of the global view. The local view shows our partner, that the number of e-mails with a zip-attachment has abruptly increased, which is an indication for an abnormality and possibly for an attack. To verify whether this is a local phenomenon, which would be speaking for a directed attack towards the local partner, the event can be compared to the global view. Doing this, we can find out, that only a few partners have recorded this phenomenon at this specific time. Therefore, the event could be a directed attack against a selected group, for example banks or insurance companies. Other partners, which have no problem at this particular moment, can use this information to protect their organizations in advanced. The centrally management Evaluation System will also be able to detect cyber war activities.

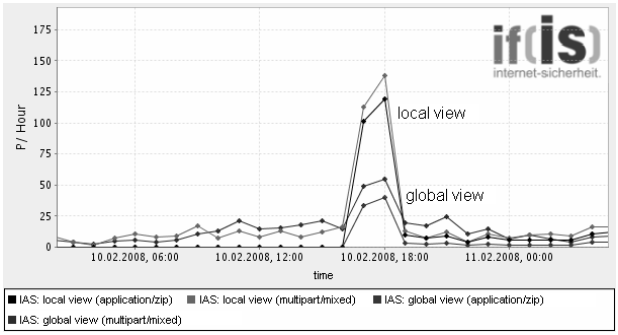

Fig. 10: example of malware detection

A further challenge will be to deal with the different natures of networks, that can be monitored with the sensor technology. The traffic passing the sensor is characteristic for the type of services,

which are provided by this different kind of network. Therefore, a content provider has a different profile of traffic as a university. To improve the outcome of the global view for these partners, they could be grouped in logical units. Each group can have their individual global view, which can be combined to the common global view of the Internet. The challenge is, that not all networks can be grouped that easily, due to their diversity.

The selection of network operators as partners should of course be done due to scientific evaluations (rectangular distribution). For the future, once the technology has reached a wider level of acceptance, this will leave some unsolved complex statistical challenges, which we have to face once this opportunity arises: What is the accurate number of sensors and where exactly do these sensors need to be placed to generate a global view in a representative manner without creating to much overhead?

In addition, we have to deal with different time zones and inconsequent daylight saving time regulations all over the world. We cannot just agree on a global time for the distributed system of sensors, because the local time reflects in the usage of services and therefore in the raw data. The traffic passing the sensor consists roughly of human and machine initiated communication. The human part is highly affected by the different time zones, as people tend to sleep during night and work during day time. The human part is also affected by cultural issues (sociological aspects) like the long lunch break in Spain (Siesta), due to the hotter climate. Besides the fact that global attacks initiated in the US by a human, might strike in Europe at night time, we also have to face the problem, that no existing real time sensor can decide for sure by analyzing the passing the communication data, if the connection was originally initiated by a human or a machine. It is not like we could build on the work of Turing, trying to test whether the communication partner is a machine or a human. The sensor is only monitoring passing traffic in real time, which in a lot of cases are only fragments of packets of the different applications.

If a segmentation of the communication into human and machine initiated would be possible, we could introduce a global time for all machine initiated communication worldwide and consider the local time for the human initiated communication. Sometimes the communication parameters give an indication, because for example some protocols are by specification for machine to machine communication. But most protocols are shared in use or might not be used in correspondence to the specification. On top of this, we have to acknowledge that the machine initiated communication is by part also influenced by the local time, since most routine jobs are run at night, to reduce the traffic load on the network. At this point further analysis is necessary. It could for instance be, that the machine initiated portion perishes in noise.

We have a strong feeling based on the results of a diploma thesis [25], that a limited but well chosen selection of parameters might already be sufficient to compare the local situation with the global overview to detect possible events. This is the great advantage of this system colleting statistical raw data, enabling the utilization of findings of this mathematical discipline. In the long run we can further extent the selected parameters, if this should become necessary. At the moment we are working on selecting and roughly splitting up the parameters. From there we can build a global view, which considers the relevant time changes for each parameter of each sensor.

3. Related work

In this 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, Data Mining algorithms, and anonymization. It also has to be analyzed on which spots of the internet the sensors need to be placed to make a representative statement [4].

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. Everybody 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 the data. Most of the input data comes from the USA and has a focus on this area, which is creating a lack of supervision on the Internet as a whole. Together with designated experts important alerts can be spoken out by the system.

As one pilot project to monitor the internet traffic in European Infrastructure, we can name the "LOBSTER" project. It is based on passive monitoring enhanced by previous experience and "*was unique in Europe and one of only three similar infrastructures that existed in the world*". [12]

The ongoing MOMENT (Monitoring and Measurement in the Next Generation Technologies) project builds besides others on the findings of LOBSTER and aims to advance the "*project towards a common and open, pan-European platform by confederating participants from various […] other measurement-related projects. The main objective is to design and implement a mediator architecture offering a unified interface for measurement services, able to use all data and functionalities from the existing measurement infrastructures. Main innovation of the project is the use of a measurement-specific ontology, allowing semantic representation and retrieval of measurement and monitoring information, as well as providing the flexibility of a service oriented architecture for future Internet applications. To validate the benefits of the integrated approach, the project will develop and demonstrate a set of tools and applications presenting the added value of combining measurement data collected from different infrastructures.*" [14]

The Worldwide Observation of Malicious Behaviors and Attach Threats (WOMBAT) project *"aims at providing new means to understand the existing and emerging threats that are targeting the Internet economy and the net citizens. To reach this goal, the proposal includes three key workpackages: (i) real time gathering of a diverse set of security related raw data, (ii) enrichment of this input by means of various analysis techniques, and (iii) root cause identification and understanding of the phenomena under scrutiny. The acquired knowledge will be shared with all interested security actors (ISPs, CERTs, security vendors, etc.), enabling them to make sound security investment decisions and to focus on the most dangerous activities first. Special care will also be devoted to impact the level of confidence of the European*

citizens in the net economy by leveraging security awareness in Europe thanks to the gained expertise." [30]

A versatile tool to collect malware is Nepenthes. "*It acts passively by emulating known vulnerabilities and downloading malware trying to exploit these vulnerabilities.*" [16] At the moment there is an ongoing project to link the Nepenthes project with the Internet Analysis System to detect and analyze the distribution of malware.

A German national initiative towards an internet early warning system is the CarmentiS project of the German CERTs in cooperation with the German Federal Office for Information Security (BSI). CarmentiS is developing a security architecture, which is supposed to be easily implemented in a straightforward fashion. Part of the project is an improved visualization and the automated detection of threats. So far the information on the German language project homepage is very limited. [1]

Other Early Warning Systems have an active access to internet services and record the availability data [10]. 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 behavior or a cyber attack. The valuation and categorization can hardly be automated. This is the point, where humans have to make decisions based on their technical knowledge, their expertise and by accessing additional information.

Another very important aspect is the distribution level for the dissemination of the warnings and information. Here, the addressees of alerts have to be selected carefully. If responsibilities and competence fields aren'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.

4. Perspective

Even if we do not know today, if we could recognize the most important attacks, we need to be able to have a Internet Situation Awareness or at least as much as the presented global view of the internet.

Similar to the situation of road traffic the results will be implemented to infrastructural security measures (Black List, Router Policies, 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. digital signature).

We face a challenging way to establish a working Internet Situation Awareness, which can be used for national or international Internet Early Warning activities. The Internet Situation Awareness will help to: (i) improve the stability and trustworthiness of the Internet, (ii) raise awareness for critical processes or components of the Internet, and (iii) find out more about the Internet and its users in order to better cater to their needs and service demands.

These aspects are getting more and more important especially in light of the convergence of fixed, mobile, circuit switched and packet based networks to all-IP based real-time capable multimedia networks. These all-IP networks try to offer the basis for citizens, enterprises, governments and organizations to perform educational, work and entertainment operations in a confidential manner with a high level of trustworthiness and availability.

To fulfill these requirements monitoring these, next generation all-IP networks and analyzing of the data-flows is essential. For this, we need to utilize new methods and techniques, algorithms and automated processing, before the open and self-governed structure of the Internet degrades, and possibly breaks down. 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 and answer interesting research questions on the "living creature" internet.

5. ACKNOWLEDGMENTS

We would like to thank Mr. Gary Warner of the University of Alabama at Birmingham, Mrs. Susanne Wetzel of the Stevens Institute of Technology and Mr. Randy Vaughn of the Baylor University for organizing this third academic conference dedicated to eCrime research. We would also like to show gratitude to the steering and the program committee. We very much appreciate the fact, that this conference gives our institute the opportunity to present this aspect of our work in the USA.

Many thanks to all the committed students, who have done their final thesis on aspects of the Internet Analysis System, and to all the research associates, who put so much of their personal time in the project.

We also want to acknowledge the trust of our partners have put into us, that have placed our sensors in their networks. We know, that our vision of a continuous Internet Situation Awareness can only be realized with the support of these partners. So many thanks to all of our partners.

6. REFERENCES

[1] CarmentiS project, website: http://www.carmentis.org/

[2] Mathias Deml: IAS Rohdaten Transfer System (*internet analysis system raw data transfer system*), University of Applied Sciences Gelsenkirchen, 2007.

[3] Stefan Dierichs: Eine strukturelle Analyse des Internets *(a structural analysis of the internet*), Diploma Thesis, University of Applied Sciences Gelsenkirchen, 2006.

[4] S. Dierichs, N. Pohlmann: "Netz-Deutschland" (*German Internet Infrastructure*), iX - Magazin für professionelle Informationstechnik, Heise-Verlag, 12/2005.

[5] Guozhu Dong, Jinyan Li: Interestingness of Discovered Association Rules in Terms of Neighborhood- Based Unexpectedness, The University of Melbourne, 1997.

[6] M. R. Endsley: Toward a theory of situation awareness in dynamic systems, Human Factors 37(1), 32-64, 1995.

[7] Michael Hahsler, Christian Buchta, Bettina Gruen and Kurt Hornik: arules: Mining Association Rules and Frequent Itemsets, http://cran.r-project.org/web/packages/arules/index.html, 2008.

[8] Uwe van Heesch: Entwicklung eines Plugin basierten Analyse-Frameworks für das Internet-Analyse-System (*development of a plugin-based analyzing framework for the Internet Analysis System*), Diploma Thesis, University of Applied Sciences Gelsenkirchen, 2006.

[9] Mika Klemettinen, Heikki Mannila, Pirjo Ronkainen, Hannu Toivonen, Inkeri Verkano: Finding Interesting Rules from Large Sets of Discovered Association Rules, 1994.

[10] Stefan Korte: Internet-Frühwarnsysteme (*internet early warning systems*), Diploma Thesis, University of Applied Sciences Gelsenkirchen, 2006.

[11] Jennifer LeClaire, Cyber Terrorism Threat Growing, EU Agency Says, CIO Today, http://www.cio-today.com/story.xhtml?story_id=1230048OPVML&nl=5, 05/27/08.

[12] LOBSTER project, website: http://www.ist-lobster.org/

[13] Piet Van Mieghem: Performance Analysis of Communications Networks and Systems, Cambridge University Press, 2006.

[14] MOMENT project, website: http://www.salzburgresearch.at/research/projects_detail_e.php?proj=127

[15] Srinivas Mukkamala, Guadalupe Janoski, Andrew Sung: Intrusion Detection: Support Vector Machines and Neural Networks, N.A.

[16] Nepenthes project, website: http://nepenthes.mwcollect.org/

[17] Binh Viet Nguyen, Self Organizing Map (SOM) for Anomaly Detection, 2002.

[18] Thomas Ostermann: Internet-Verfügbarkeitssystem (*internet availability system*), Diploma Thesis, University of Applied Sciences Gelsenkirchen, 2006.

[19] Romualdo Pastor-Satorras, Alessandro Vespignani: Evolution and Structure of the Internet, A statistical physics approach, Cambridge University Press, 2004.

[20] 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.

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

[22] N. Pohlmann: "Internetstatistik" (*statistics of the internet*), Proceedings of CIP Europe Publisher, B.M. Hämmerli, 2005.

[23] Marcus Proest: Die globale Sicht auf das Internet (*the global view of the internet*), University of Applied Sciences Gelsenkirchen, 2005.

[24] Marcus Proest: Entwicklung einer Sonde für ein Internet-Analyse System (*development of a sensor for an internet analysis system*), Diploma Thesis, University of Applied Sciences Gelsenkirchen, 2005.

[25] Gianfranco Ricci, Betrachtung der vom IAS gesammelten Kommunikationsparameter auf Relevanz zur Anomalie und Angriffserkennung (*evaluation of the relevance for the detection of abnormalities and attacks of the communication parameters collected by the internet analysis system*), Diploma Thesis, University of Applied Sciences Gelsenkirchen, 2008.

[26] Sebastian Spooren, Entwicklung eines profilgestützten Visualisierungssystems zur Darstellung von raum- & zeitbezogenen Soll-/Ist-Abweichungen (*development of a visualization tool for the IAS*), Diploma Thesis, University of Applied Sciences Gelsenkirchen, 2007.

[27] Marina Thottan, Cuanyi Ji: Anomaly Detection in IP Networks, 2003.

[28] Sven Tschöltsch, Konzeption und Realisierung einer globalen Sichtweise auf das Internet zur Bewertung der eigenen Sicherheit (*concept and realization of a global view of the internet for a better evaluation of the local security situation*), Diploma Thesis, University of Applied Sciences Gelsenkirchen, 2008.

[29] Svenja Wendler: Entwicklung eines Anaylsemoduls zum Internet-Analyse-System – Finden von Strukturen im Internetverkehr in Form von Assoziationsregeln (*development of an analyzing module for the internet analysis system – discovery of structures in the internet traffic consisting of association rules*), Diploma Thesis, University of Applied Sciences Gelsenkirchen, 2006.

[30] WOMBAT project, website: *http://wombat-project.eu/*