Objectives and added value of an Internet Key Figure System for Germany

Sebastian Feld · Tim Perrei · Norbert Pohlmann · Matthias Schupp

Institute for Internet Security
Gelsenkirchen University of Applied Sciences
{feld | perrei | pohlmann | schupp}@internet-sicherheit.de

Abstract

This work is motivated by the fact that the Internet can be seen as a critical infrastructure, whose on-going operation is particularly worth protecting. Problematic when considering the state of the Internet are two things: On the one hand, there are many dependencies in the context of the Internet, on the other hand there are only a few key figures that allow comprehensive statements. In the course of this work, a Key Figure System will be described in which the control object is the Internet by itself. The complex structure of the Internet is to be made more transparent and the condition, changes and future potential are to be expressed. In addition to the various objectives that are to be achieved during the design and implementation of such an Internet Key Figure System, this work describes the problems that have to be solved. These are less technical, but in fact organisational and legal nature. Each Key Figure System requires a control object, that is a clearly defined scope, which the data collection, data processing and data visualisation refer to. In the following the definition of an „Internet Germany" is given and the appropriate stakeholder and criteria are described. This work concludes with an explanation of the different added value of an Internet Key Figure System for the various addressees.

1 Motivation

Today's Internet consists of a variety of networks, called Autonomous Systems (AS), which are operated by Internet Service Providers (ISPs), large companies and universities. There are currently more than 37,000 AS[1], which build the Internet with more than 70,000 connections. For a more detailed consideration and a proper assessment of the significance of each AS, it is important to see what role each AS occupies in the interaction of the Internet. AS are acting completely independent, so the operators have different strategies on how to organise the communication of IP packets on their network with the help of routing protocols.

Not only within Germany, the Internet as a part of the information and telecommunications technology is reckoned to the critical infrastructures [BMI09]. The impairment or loss of parts of the Internet can have an enormous impact. For example, the disturbance of IP telephony as the Skype outage in 2007[2] can lead to sustainable economic damages if a company is no longer able to conduct telephone business.

1 See, amongst others, the routing table of the project "Route Views" under [Rout11].
2 See [Skyp07].

N. Pohlmann, H. Reimer, W. Schneider (Editors): Securing Electronic Business Processes, Vieweg (2011), 1-11

Currently, an on-going operation of the Internet is essential. To ensure the most trouble-free operation, it is necessary to be able to observe the current state and also to estimate the future development of the Internet. It is the only way to face novel events (both positive and negative) optimally.

Basically the motivation for an Internet Key Figure System can be divided in two aspects: First, it is about the many different dependencies in the context of the Internet and second, it is about the barely available, pre-existing key figures.

1.1 Dependencies in the context of the Internet

There are many different dependencies in the context of the Internet, and in some cases, a dependency on a country – often the United States of America – can be noted. The dependencies can occur at various levels, including:

- Technical dependencies, such as the combination of parts of the Internet by means of intercontinental-laid undersea cable
- Dependencies at the service level, for example the „web surfing“ without the transparent use of the Domain Name System (DNS) is hardly feasible
- Administrative dependencies, for example the Internet Corporation for Assigned Names and Numbers (ICANN) coordinates, among other things, the management of top-level domains

In the following two examples of dependencies in the context of the Internet are highlighted.

There are few very large AS (so-called Tier 1 providers) that connect large parts of the Internet and thus achieve a connectivity of all end systems all over the Internet. These are enormously important for the stability of the Internet. Currently the largest and most important AS are American. The largest German AS, the one of Deutsche Telekom, can be found in the lower places of the TOP25 sorted by the number of connections[3].

Another example are Border Gateway Protocol (BGP) router, which are necessary for the smooth operation of the Internet. Under certain circumstances, the failure of particular BGP router can cause many user of the Internet to be unreachable. These may be customers of an ISP or an entire nation[4].

The idea of introducing an Internet Key Figure System is to make the complex architecture of the Internet more transparent and to express its condition, changes and future potential. The Internet economy receives a common Internet Key Figure System, with which the current state of the Internet can be represented with respect to different scales (see section „5.2 Stakeholder and criteria“).

1.2 Barely statistical key figures

There are barely statistical key figures for the critical infrastructure Internet. Though there are local data silos, these are not in a broader context.

3 See, amongst others, the routing table of the project “Route Views” under [Rout11].

4 Libya can serve as an example, which was completely separated from the rest of the Internet for some time in the spring of 2011. See, amongst others, [Heis11].

For example, large website operator or services for web traffic analysis (such as „Alexa Internet“ [Alex11]) can make statements about the distribution of the website visitors' used operation systems, web browsers, software and the like. E-mail provider or blacklist operators can provide statistics on the current amount of spam. Additionally, projects such as „Route Views“ [Rout11], BGPmon [Bgpm11] or „RIPE Atlas“ [Ripe11a] provide information regarding the connection of Autonomous Systems or the availability of Internet services.

The various information on aspects of the Internet offer, due to the lack of a global or comprehensive nature, only a limited statement. Many findings can only be generated when different data are linked. For example, a message about a serious vulnerability in a web browser is more relevant if the software is actually used by many users. If there would be this global or at least "higher" perspective, one could measure the current state of the Internet, assess the development of the Internet better and thus make better-informed decisions for the future.

2 Basic idea of an Internet Key Figure System

Key Figure Systems are used in the area of business administration for the quick receipt of concentrated information about a company's performance and efficiency. They can also assist in planning, monitoring and controlling of a company.

Such a **Key Figure System** describes an ordered set of interdependent business key figures. It aims to inform as fully as possible about a given situation. Thus a Key Figure System groups and processes logically related key figures. Thereby the information content of the Key Figure System is about to be higher than the sum of the information content of the individual key figures.

Basically a **key figure** describes an exact quantifiable measure, which results from a reproducible measurement of a parameter, state or process. In the context of a Key Figure System a comprehensible collection of the key figures is most important. Key figures can be divided into two areas: Absolute (atomic) key figures and relative key figures. Absolute key figures are those that are integrated from outside into the Key Figure System and can not be disaggregated further. Relative key figures, however, arise from the relationship of other key figures and are formed within the Key Figure System.

An **Internet Key Figure System** is in conclusion a Key Figure System in which the situation subject to investigation is not a company, but the Internet itself. It collects key figures that relate to the Internet or that are generated by the Internet.

The idea of introducing an Internet Key Figure System is to make the complex structure of the Internet more transparent and to express its condition, changes and future potential. The Internet economy receives a common Internet Key Figure System, with which both the current state of the Internet can be represented and a retrospective consideration with respect to the measured scales can be carried out. In the course of this work the objectives, problems and added value in the design, implementation and use of an Internet Key Figure System will be treated in detail.

3 Objectives in the design of an Internet Key Figure System

An Internet Key Figure System has to provide a view on the Internet as comprehensive as possible. This requires many different aspects of the Internet to be treated and also the individual findings to be linked.

There are a number of objectives that an Internet Key Figure System has to reach. They can be divided in the three following components: Data Collection, Data Processing and Data Visualisation.

3.1 Data Collection

An Internet Key Figure System should be able to read **various data sources** in the actual data collection component. It is also important in terms of the significance of the gathered data to collect data as general as possible. For example, active measurements must take geographical backgrounds and different technologies into account. Availability and quality measurements must take place from different locations in a wide-ranging way and at best via various access technologies as well.

The data sources also have different requirements regarding the **refresh period of the key figures** that must be met through the Internet Key Figure System. We divide the requirements in three different time periods. The shortest time range covers minutes and hours. Key figures with this requirement may already have a significant amount of impact on the Internet Key Figure System just with short-term changes. In particular, data on availability of services on the Internet are very important. The medium time range covers days and months. Key figures are sorted here, when changes show medium-term effects on the Internet Key Figure System. This can include routing data from the Boarder Gateway Protocol. For example, the degree of "intermeshing" of the various Autonomous Systems can thus be observed. The third time period is for long-term data, which change only rarely or over extended periods of time. A survey is meant to be in the area of quarters or years. Interesting data are those about the development of the Internet's infrastructure, such as the average end connection's bandwidth.

Based on the different refresh periods, **various measurement methods** arise. Short and medium-term data updates are of practical use only through automated procedures. Therefore, the Internet Key Figure System must be able to independently collect data using external sensors or systems. Datasets, that have to be collected over long periods, need generally to be applied to manually.

In order to compare different key figures and to interpret the collected information, there has to be defined a certain **scope**. This may be the global Internet or a subarea such as an "Internet Germany" (see section 5.1 "Definition Internet Germany").

In connection with the scope of the Internet Key Figure System is the definition of **criteria**. It is determined which realms or layers of the Internet will be measured and treated. Defined criteria roughly structure an Internet Key Figure System (see section 5.2 "Stakeholder and criteria").

3.2 Data Processing

An Internet Key Figure System requires an **analysis and evaluation module**, with which the information collected can be processed. Different algorithms for data analysis, especially those in the field of data mining, are needed. In addition, methods are required to describe the "normal state" of the measured part of the Internet. Building on this, techniques for anomaly detection are necessary in order to detect changes in the key figures automatically. Not only the current state, but also the development of the key figures should be estimated and evaluated.

Additionally, an Internet Key Figure System should be able to highlight **different relationships** between the key figures. This can be for example logical relationships ("threat potential = threat / utilisation ratio", "encrypted = total - unencrypted"), empirical relationships ("greater use of firefox browser causes more AES/SHA1 with SSL") or hierarchical relationships ("more botnet activity leads to more DDoS attacks leads to more SYN requests").

Finally, the information gathered need to be stored in the **backend** of the Internet Key Figure System properly. Under certain circumstances there can be accumulated very much information. This is why a mechanism is needed that stores alongside established key figures also significant detail data in a privacy-compliant and economical way. This allows a retrospective analysis for relationships that were not considered or known in advance – that means at the time of data collection.

3.3 Data Visualisation

An Internet Key Figure System must be in the position to represent the collected and consolidated data specifically. There are various tools needed to allow **different views** on the key figures. Expert tools for a detailed analysis allow the user to display certain key figures (for example as time series or pie chart) and to compare them. It must be easy to verify assumptions about relationships (see section 3.2 "Data Processing") between different key figures. An information portal provides a general overview, without going into the depths. Real-time relevant information can be reduced to the essentials and be made available for example as barometers. Finally, an Internet Key Figure System must provide reporting functions. A user should be able to generate comprehensive reports on arbitrary time periods and datasets, in order to perform a quick survey or recapitulation of past developments, for example. All options for visualising the information should be designed for different devices, including implementations as a web application or "app".

4 Problems with the development of an Internet Key Figure System

Creating a Key Figure System described above is extensive and comprises many tasks and problems to be solved. Most important are less the technical aspects, but the organisational and legal aspects.

Regarding the technology, the implementation of an Internet Key Figure System is indeed a challenging but very solvable problem. During data collection, interfaces are necessary for the integration of third party data as well as sensor technologies for the active measurement of certain

values. Within data processing, various algorithm for data analysis, especially those in the field of data mining, are needed. Building on this, techniques for anomaly detection are required in order to detect changes in the key figures automatically. The data visualisation component can access many frameworks that can represent comprehensive information intelligently.

In the following four problem areas are presented, which has to be solved during the development and operation of an Internet Key Figure System.

4.1 Participating companies

One of the biggest problems is to identify and to integrate participating companies. Like any other key figure system an Internet Key Figure System also relies on a variety of data. Some information can usually be collected, processed and visualised with no effort by the operator of the system itself. There are also lots of data, which have a considerable explanatory power, but are not freely accessible. A combination of these data with existing information would possibly bring new findings and thereby constitute an added value (see chapter 6.1 “Identification of relationships”).

Without corresponding agreements, it is hardly possible to incorporate non-public information in the Internet Key Figure System. This is particularly true if the company is not interested in the results of the Key Figure System or does not see any added value in the cooperation. Does the operator of an Internet Key Figure System fails to motivate relevant enterprises into a cooperation or to convince them of the added value, the development and operation of an Internet Key Figure System turns out to be difficult.

4.2 Privacy

Another focus is on the privacy. The required data must be collected and stored in a manner that meets two requirements.

On the one hand, the necessary privacy of each collected information must be met. There are data that are less critical – in the sense of privacy – than others. As an example, the number of DSL lines in Germany at a certain time shall be compared with the communication parameters measured at an Internet exchange point within a given period.

On the other hand, there has to be created a solid date base in spite of the privacy-compliance, which allows a later reinterpretation of the data. The retrospective interpretation of a dataset regarding criteria, that were not considered in a first analysis, must be feasible.

4.3 Scope

It is necessary to define a clear demarcation for the measured part of the Internet (a scope). This has the background that the key figures and the derived findings must be interpretable and comparable. For example, the demarcation of the Internet can be accomplished by a country (see section 5 “Internet Key Figures for Germany”). Due to the nature of the Internet, it is difficult to draw clear “boarders”. The significance of some key figures is strongly influenced by the definition of the scope, for which reason a certain fuzziness of the used Internet Key Figure System remains.

4.4 Financing

Another problem is the financing of an Internet Key Figure System. A Key Figure System as such does not achieve any revenue from operating activities. It rather relies on the fact that interested groups and companies, that possibly are involved in data collection and data processing, involve in maintaining the Key Figure System. At this point it is again the task of the Key Figure System's operator to create incentives for using and supporting the system.

5 Internet Key Figures for Germany

The necessity for an Internet Key Figure System described in this work is also specifically given for the Federal Republic of Germany. A system is needed that allows to collect and process key figures for the Internet. Such information, made available and compressed, about a to be defined "Internet Germany" (see section 5.1 "Definition Internet Germany") can serve appropriate points as a basis for the creation of options for actions regarding the Internet. Targets of these recommendations may be different decision-maker: The politics, participating companies, or even citizens. Using an established Internet Key Figure System it is possible to evaluate, whether implemented actions achieved the desired effect.

5.1 Definition "Internet Germany"

One of the most important requirements for a Key Figure System for Germany is the definition of a scope. One way to define an "Internet Germany" refers to the Autonomous Systems (see figure 1):

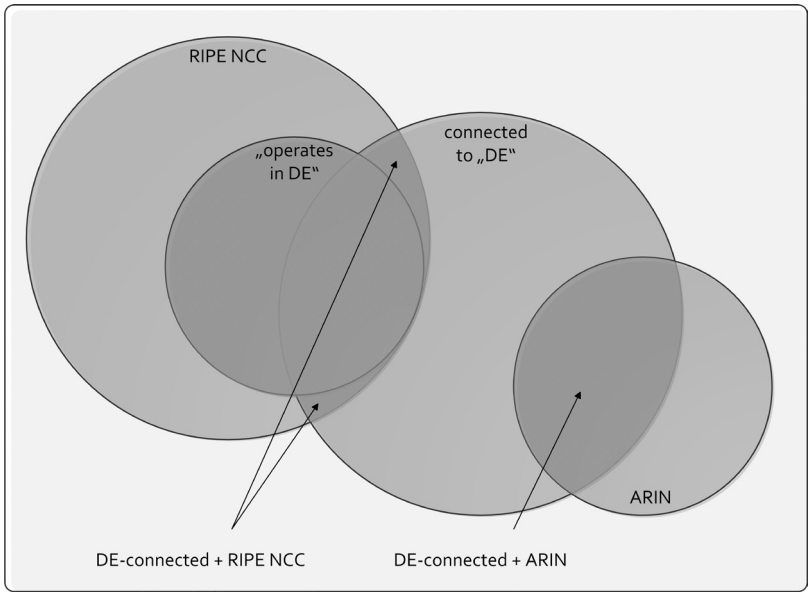

Fig. 1: Possible definition of an "Internet Germany" via Autonomous Systems.

From the approximately 37,000 currently active Autonomous Systems on the Internet only those are relevant, where the operator is active in Germany. This can be compared, for example, on the

geographic use of the assigned IP addresses (for instance via a GeoIP database), or the country classification of the RIPE database [Ripe11b]. Beyond that, such Autonomous Systems are counted to the "Internet Germany", which are directly connected with the aforementioned AS and in addition are placed in Europe or North America. This corresponds to the registration at RIPE NCC or ARIN.

5.2 Stakeholder and criteria

The Internet in general – and thus the "Internet Germany" in particular – can be divided in different aspects or actors, which contain all the elements involved in the Internet. In business administration these views on a control object are also called perspectives [Kuet03].

The Internet's aspects are divided into four areas:

- **Participants**: All Internet-connected user end devices (PCs, laptops, smart phones, cars, refrigerators, …).
- **Infrastructure**: Equipment and structures, which allow participants to use the Internet's services (AS with BGP routers, connections, …).
- **Services**: Any provision of service that are made available to the Internet's users (e-mail, Web, VoIP, …).
- **Threats**: These are all dangers lurking on the Internet for other aspects of the Internet (malware, botnets, …).

From the four aspects those criteria are derived, about which the Internet Key Figure System has to provide information:

- **Capability**: Represents the infrastructure aspect of the Key Figure System and contains parameter that make statements about the capacity and dependencies of the Internet. Such parameters include the average hop count, bandwidth and packet loss rate. Furthermore, the information is relevant, what proportion of the Autonomous Systems can be reached without the use of transit.
- **Availability**: Represents the services aspect of the Key Figure System and covers statements regarding the availability of services on the Internet from the perspective of end users.
- **Usage estimation**: Represents the participants aspect of the Key Figure System and contains parameter that make statements about the utilisation of the Internet and the technologies used. These include, for example, the penetration rates of operating systems, browser software and access technologies such as DSL or UMTS, as well as the breakdown of the total data volume.
- **Threat potential**: Represents the threats aspect of the Key Figure System and covers parameter that make statements about the threat of the Internet. Examples for a threat indicator can be the number of new virus signatures per measured period, the total number of infected websites or the measured data rates of DDoS attacks.

Using an Internet Key Figure System various key figures with regard to different criteria of the "Internet Germany" are made available. Via these key figures, a comprehensive view on the German Internet in terms of different aspects are getting possible.

6 Added value through the use of an Internet Key Figure System

The objective of an Internet Key Figure System for Germany is to provide added value, which would not be possible without such a system. The first of the two most important added value is the possibility to detect relationships between different measurements. This is possible at all by the second added value, the comprehensive context of the key figures. Other added value are the possibility to provide the Internet Key Figure System as a data base to third parties, as well as the supply of a comprehensive data base for the generation of options for actions.

6.1 Identification of relationships

Probably the biggest added value of a Key Figure System for the Internet is the collection and consolidation of much information that promote the understanding of the critical infrastructure Internet. For example, availability key figures can be determined to give information about the accessibility of websites and services. Statistics on protocols and technologies of the Internet are possible if anonymous data of central Internet infrastructure points is collected and processed. Interesting aspects are, for example, the distribution of used web browsers, operating systems, or the ratio of IPv4 to IPv6. It can be given a view of the interdependence of the Internet, if (freely available) BGP routing data is retrieved and processed. In particular, the connections of the AS among each other are interesting. Different service provider on the Internet have massive data bases with respect to the Internet and make parts of it freely accessible for the retrieval via APIs. New findings by a combination and analysis of the key figures become possible. Infrastructure data, such as information on the spread of mobile devices, Internet access technologies and more, can be integrated by statistics or reports of various institutions. Publicly available vulnerability databases of security initiatives and security barometers can be used in order to collect the spread of security breaches, damage due to cyber crime or spam volumes and more.

6.2 Comprehensive Context

Through a comprehensive Key Figure System there is a data base available that not only exists locally, but has also a broader context. Knowledge can be generated, the current state of the Internet can be measured, the development of the Internet can better be assessed and thus better-informed decisions for the future can be made. In summary, possibly unknown relationships can be better established, forecasts using the newly acquired findings be created and developments be explained.

6.3 Data base for third parties

A recently described Internet Key Figure System will collect much information and perform a lot of analysis and interpretations, but also provides a good basis for further work and observations by third parties. On the one hand a collaboration is thus possible through participating companies together with their key figures provided. On the other hand, the system can be used by third parties and accordingly serve as a data base itself. Users can not only base on experience, but also refer to specific numbers and draw their own conclusions.

6.4 Generation of options for actions

An Internet Key Figure System provides plenty of processed and compressed information, which can be used as a base by competent authorities to create options for actions. Targets of these recommendations are different decision-maker: The politics, participating companies or citizens. In addition, an evaluation is possible whether the measures implemented have achieved the desired effect.

- **Politics**: Good information is the basis for effective management and proper decisions on regulatory and legal issues. Statements can be confirmed using the Internet Key Figure System where possible or arguments can be disproved.
- **Participating companies**: Good information can reveal trends and developments in the Internet market and thus improve the basis for strategic decisions. In addition, internal company data can be correlated with Internet Key Figures and thus be enriched or cross checked.
- **Citizens**: Appropriate represented information can promote the interest in sub topics of the Internet – for instance the aspect of security. Awareness raising can take place, for example, when a strong increase in phishing emails or infected websites are presented and appropriate warnings or behaviours are issued.

7 Summary

An Internet Key Figure System that meets the requirements described in this work is a powerful expert tool for performance control and trend detection on the Internet. In particular, a specialised focus on a certain part of the Internet, for example an "Internet Germany", allows findings that would not be that apparent within generalised (global) systems. Furthermore, the current state of the Internet can be measured, normal conditions can be defined and deviations or anomalies can rapidly be detected.

The collected data allow, by means of their quantification, to discover, substantiate and analyse possibly existent, but yet unknown relationships. The type of key figures, relationships and statements is not limited to one particular aspect – such as an exclusively economic or technical issue –, but depends largely on the rules that generated and analysed the key figures. The composition and analysis of the key figures provide a great potential for further research and work.

Far from that, the success of an Internet Key Figure System depends mainly on the participating companies. Like any Key Figure System an Internet Key Figure System is also only as good as the data entered. In addition to the freely available data, the on-going and voluntary contribution of so far unpublished data is absolutely desirable and necessary. This enhances the significance of the Internet Key Figure System.

Acknowledgment

This work is part of the project "Deutscher Internet-Index (DIX)"[5], which is funded by the Federal Ministry of Economics and Technology (BMWi). The contents of this publication is solely in charge of the authors and reflect in no way the BMWI's opinion.

5 German for „German Internet Index".

References

[Alex11] Alexa Internet Inc.: Alexa the Web Information Company. http://www.alexa.com (Last Access: 01.08.11), 2011.

[Bgpm11] BGPmon: BGPmon.net, a BGP monitoring and analyzer tool. http://bgpmon.net (Last Access: 01.08.11), 2011.

[BMI09] Bundesministerium des Innern: Nationale Strategie zum Schutz Kritischer Infrastrukturen (KRITIS-Strategie). Berlin, 12. Juni 2009.

[Heis11] heise Netze: Internet-Abschaltung: Libyen hat von Ägypten gelernt. http://www.heise.de/netze/meldung/Internet-Abschaltung-Libyen-hat-von-Aegypten-gelernt-1206016.html (Last Access: 01.08.11), 2011.

[Kuet03] Martin Kütz: Kennzahlen in der IT - Werkzeuge für das Controlling und Management. dpunkt Verlag, 2003.

[Ripe11a] RIPE Network Coordination Centre: RIPE Atlas. http://atlas.ripe.net/ (Last Access: 01.08.11), 2011.

[Ripe11b] RIPE Network Coordination Centre: Delegated RIPE NCC Latest. ftp://ftp.ripe.net/pub/stats/ripencc/delegated-ripencc-latest (Last Access: 01.08.11), 2011.

[Rout11] University of Oregon Route Views Project: Route Views Project Page. http://routeviews.org (Last Access: 01.08.11), 2011.

[Skyp07] Skype Technologies SA: What happened on August 16. http://heartbeat.skype.com/2007/08/ (Last Access: 01.08.11), 2007.