CarmentiS: A Co-Operative Approach Towards Situation Awareness and Early Warning for the Internet

Bernd Grobauer[1], Jens Ingo Mehlau[1], and Jürgen Sander[2]

[1] Siemens AG, Corporate Technology IC CERT, Otto-Hahn-Ring 6, D-81730 München, bernd.grobauer|jens.mehlau@siemens.com,
[2] PRESECURE Consulting GmbH, Beelertstiege 2, D-48143 Münster, js@pre-secure.de

Abstract. Although plenty of organizations collect sensor data such as IDS alerts or darknet flows, local analysis has its definite limits when it comes to derive conclusions about happenings and trends within the Internet as a whole.
CarmentiS, a joint effort of the early warning working group within the German CERT association, provides an infrastructure and organizational framework for sharing, correlating and cooperatively analyzing sensor data. The infrastructure allows organizations to submit sensor data – at the moment, net flows and IDS alerts are treated – over a secure channel to a central database. Cooperative analysis of the data is made possible via a secure web front end allowing analysts of participating CERTs to create and execute analysis profiles as well as share and discuss analysis results. Thus correlating sensor data and pooling know how and resources for analysis from different sites, CarmentiS provides a framework for a co-operative approach towards situation awareness and early warning for the Internet.
This article gives an overview of the CarmentiS infrastructure and organizational framework, and describes the current status of the project. It also addresses open questions that can only be solved by experimenting with co-operative analysis and gives an outlook of possible further developments of the CarmentiS approach towards improved situation awareness and early warning.

1 Motivation

CERTs must recognize new threats to the IT infrastructure of their constituency as early as possible and take appropriate counter measures to avoid or at least limit damage. Although many teams collect and analyze sensor data such as IDS alerts or traffic information extracted from the logs of network components, sensor data usually plays a minor role as far as *early* warning is concerned. The reasons are at least twofold. First, although indications of a new threat such as tell-tale signs of a new attack pattern may occur on one's local network, it is far more likely that early-warning indicators appear first on other networks.

Secondly, even if there are early-warning indicators to be found in local sensor data, analysis capacities are often limited and likely to be taken up with the handling of more clear-cut incidents. In the cases where a more explorative form of analysis is performed, often outside help from other teams is sought usually with the first step of finding peers who have observed a similar phenomenon on their network. Projects like DShield [1] and MyNetWatchman [2], which correlate basic sensor data, namely firewall logs, contributed by voluntaries, offer only very limited possibilities for such correlation of locally observed phenomenons with more wide-spread occurrences, so – somewhat ironically – mailing lists are often still the medium of choice to interact with peer analysts.

Thus, at present we have a situation in which plenty of network data is being collected locally from many teams but used mostly for gaining situation awareness over the local network. There are correlation projects attempting to create a more global view, but with a restricted scope. First of all, existing projects operate only with one kind of data. For example: the aforementioned DShield and MyNetWatchman operate with firewall logs, which in most cases means that connection attempts to blocked ports are being logged; eCSIRT.net [3] collects and correlates IDS alerts; the IMS project [4] analyzes darknet traffic. What is more, none of the existing projects provides capabilities for joint analysis of collected data, even though joint analysis – pooling both analysis resources and analysis know-how from different teams – has in the past again and again proved invaluable for interpreting newly discovered phenomenons.

CarmentiS, a joint effort of the early warning working group within the German CERT association with support by the German Federal Office of Information Security (BSI [5]), attempts to provide an infrastructure and organizational framework for sharing, correlating and cooperatively analyzing sensor data of several kinds. With the CarmentiS framework, local situation awareness of single teams can be extended towards global situation awareness, a necessary basis for building up early warning capabilities. Section 2 of this article provides a closer overview of the CarmentiS approach to situation awareness and early warning, Section 3 informs about the current development status, and Section 4 provides an outlook on the next steps in the CarmentiS project.

2 CarmentiS Approach

Pursuing a cooperative approach for building an early warning system, one has to bring together different teams of different organizations. Challenges to do so lie both in supplying an appropriate technical infrastructure that supports cooperation as well as defining an adequate organizational framework. The following section describes the main participants of CarmentiS that have been identified so far and the CarmentiS architecture to bring these participants together.

2.1 Participants

In a first step, three types of stakeholders are identified and supported by CarmentiS:

– *Partners:* Partners represent organizations, which deliver data of interest towards the CarmentiS central. Rules and regulations regarding the use of the data and analysis results have to be established between the partners an the host of CarmentiS. Each partner has to accept these rules for the data delivered to the CarmentiS central. In other words, it is each partner's responsibility to assure that the delivered data may indeed be exported to CarmentiS.
  A second important role for partners is the provision of resources and capabilities for cooperative analysis (although there may be partners that only deliver data). Usually, these capabilities will be provided by a partner's CERT team.
– *CERTs:* Analysis results and early warning information created by co-operative analysis within CarmentiS are not only of interest to the CarmentiS Partners' CERTs, but also to other CERTs: in most cases, an organization's CERT is the ideal contact for delivering information and warnings relevant for that organizations IT security. Therefore, CarmentiS envisions CERTs that for some reason cannot act as CarmentiS partners as ideal recipients for informations and warnings concerning their constituency.
– *Governance / CIIP:* Critical Information Infrastructure Protection (CIIP) is a main task for national governance systems. Protecting critical infrastructures, such as communications, transportation, and energy, against disruption of any kind is increasingly crucial in maintaining both domestic stability and national security. By providing situation awareness and early warning, CarmentiS will be able to support approaches towards improved critical information infrastructure protection based on early warning as.

2.2 Architecture

The cooperative approach of CarmentiS is based on the following simple idea: organizations have situation awareness of their own networks, but knowledge of what is going on behind their perimeters is often missing. In order to broaden the range of vision, participants deliver different types of data of interest to an independent third party. This intermediary, named CarmentiS central, provides the main functionalities for receiving data from partners, conducting analyzes of this data, and presenting appropriate user functions for analysts as well such as CERTs and CIIP-related users. It consists of four main components: the *Import Interface and Storage* component, the *Main Analyze Component*, the *Analysts Workbench*, and the *User Workbench* (see Figure 1).The following sections describe these parts of the CarmentiS architecture including the dynamic behavior of the data export process.

Import Interface and Storage

The two main issues regarding data import are data volume and privacy concerns. Depending on characteristics and placement of the deployed sensors, very

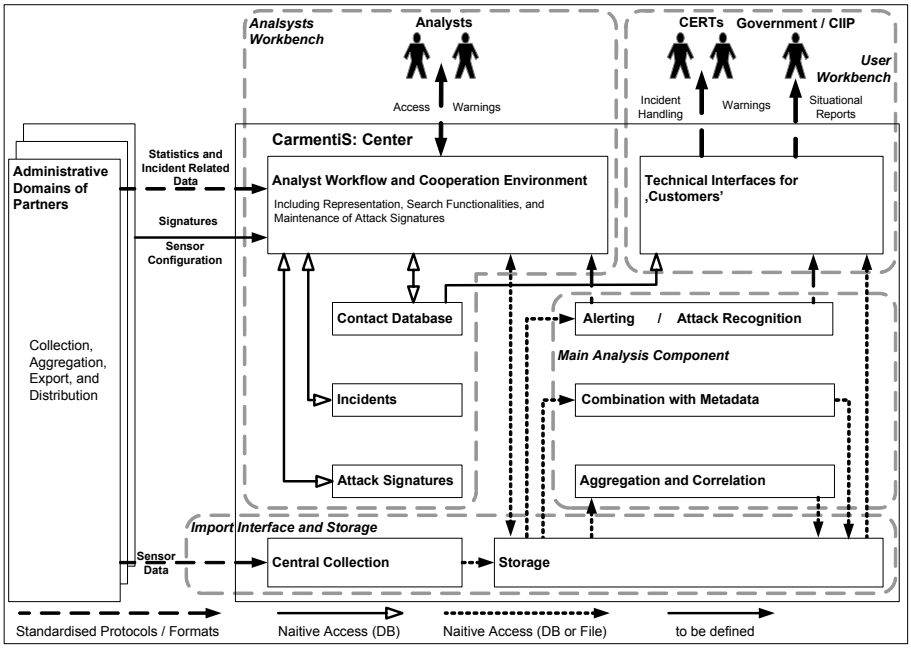

Fig. 1. CarmentiS Architecture

large data sets may be generated. Much of this data is likely to be of a sensitive nature, either because of data-protection laws or because information could be gathered from the data that is considered confidential by the institution whose traffic is being monitored. It is to be expected that CarmentiS partners differ in their assessment of what should be exported and what should not. Clearly, for correlation and analysis, receiving as much information as possible is advantageous. Therefore, predefining a fixed import policy that meats the least common denominator between the CarmentiS partners would be counterproductive. Instead, CarmentiS import mechanism should supply a policy controlled export that allows each partner to specify without much effort which data should be exported and how much anonymization/pseudonymization is to be exercised: as pointed out in Section 2.1, each partner is responsible for ensuring that the data delivered to CarmentiS may indeed be shared.

Information about the export policy that was applied to the data is one important aspect of meta information that must accompany all submitted data; other examples of meta information are sensor configuration, sensor location, etc. Without such meta information, sensible interpretation and correlation of data is impossible. The CarmentiS data exchange format therefore must support the communication of meta data.

Because of the sensitive nature of the transmitted data, the transmission channels to the CarmentiS central have to be secured using state-of-the-art authentication and encryption mechanisms.

Within CarmentiS central, the component *Central Collection* receives the data send by the partner and removes the envelopes of the encapsulated datafiles. After that, the extracted data are stored in the main *Storage* for further analysis. Because of the nature of the delivered data, mechanisms have to be found that can deal with very large data sets.

Data Analysis

The main task of the *Data Analysis* component is to aggregate and correlate the data delivered from the partners, to conduct analyzes and to give alerts.

- *Correlation of different data* Because various kinds of data from different organizations are collected, the data has to be aggregated in an appropriate way. Correlation measures are needed for data resulting from NIDS and Netflows; the integration of additional kinds of data will require additional correlation mechanisms.
- *Profile-based analysis* CarmentiS has chosen profile-based analysis as an appropriate mechanism for cooperative analysis. Profiles are developed and dynamically updated by analysts. They are used by the partner for analyzing the data. Roughly speaking, the analysts profiles capture the expert knowledge of intrusion detection specialists; thus, expert knowledge about analyzing sensor data can be exchanged between analysts of different teams. The analyzes may be based on the overall data of CarmentiS, data of a single partner, or on the aggregated data of specific groups of interest. The latter approach could provide a possibility to examine specific sectors of critical infrastructures by grouping partners into their respective sectors. By conducting the same analysis on different input data, one can gather additional information by comparing the findings.
- *Automated analysis* Complementing profile-based analysis carried out by analysts, proven automated analysis methods will be necessary to support the analysts, e.g., by creating notifications about events of interest that warrant closer analysis.
- *Alarm notification* Automatically generated warnings and in some cases also notifications should be distributed using a push rather than a pull model to ensure timely response. In order to further improve response times, the such messages must be based on communication standards such as IODEF [6] thus facilitating an import to standard incident response tools like SIRIOS [7].

Analysts Workbench

The left part of the CarmentiS central in Figure 1 depicts the components necessary for the analysts workbench. In order to provide a cooperative analyze, it is necessary to build virtual teams of analysts, which are employees of

the participating partners. The cooperation of the analysts is coordinated and supported by the analysts workbench as follows:

- Presentation of information describing the actual overall security status, analyzes, and technical as well as non-technical indicators for possible malicious activities.
- Presentation of the findings of the analyzes conducted by other analysts regarding the danger of attacks.
- Presentation of reports describing well-known as well as upcoming attack techniques. This provides valuable background information for finding new attacks, accurately adjusting the CarmentiS sensors, and designing appropriate countermeasures.
- Providing an interface, which enables the analyst to develop methods or countermeasures for identifying and combating new attacks.
- Providing capabilities for distributing warnings and advisories via E-Mail or SMS.
- Providing an interface for adjusting CarmentiS sensors. This includes capabilities for directly update sensors placed at the partners as well as providing new signatures for download.

In detail, the analysts workbench administrate the signatures and require access to a contact database, an incident database, and an knowledge database.

User Workbench

The findings of the analyzes are presented to CarmentiS users via a web-portal. The user interface should support different views specialized for each participating group of stakeholders. Warnings are sent using a push model (e.g., email) but may very well be duplicated within the the user workbench to provide a comprehensive overview.

3 Development and Status

This Section gives an overview of the development status of CarmentiS components for importing, storing and analyzing sensor data. Of the components described in the previous section, the export component (placed at each partner's site) and the data storage unit have been implemented completely. With the extension of an existing open source tool for the analysis of Netflows to cover also NIDS data and meta data, a powerful tool for data analysis has been integrated into the existing infrastructure. The user interface of this analysis tool forms the first component of the analysts' workbench.

3.1 Data Import

Figure 3.1 shows the data export work flow, as implemented by most of the CarmentiS partners. An export tool was developed within the CarmentiS framework with the following features:

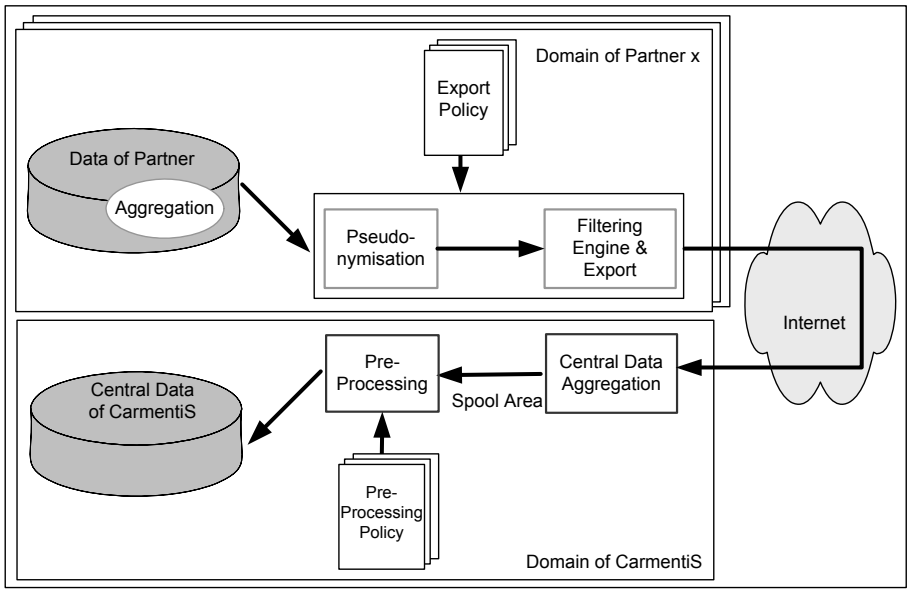

Fig. 2. Data Export

- *Support of various data formats* The export tools supports up to now three main input formats, namely CISCO Netflows, Argus and IDMEF [8]. The design of the tool is modular, therefore additional formats can be simply integrated. All data will be delivered in a common format specified within the framework. The common format is not only a container for existing data formats, but provides meta information needed for proper analysis, e.g. class of information, kind of sensor, time frame of the data, location of the sensor, etc.
- *Filtering-Engine* This module can be used to filter the captured data according to a given policy. In order to fulfill the requirements of an administrative domain, it is possible to drop any connection or event based on network, IP address, port or protocol.
- *Pseudonymization* Within a single administrative domain, pseudonymization of IP addresses is simple: methods such as CryptoPan [9] provide a mapping from real IPs to pseudonymized IPs. For pseudonymization within the CarmentiS framework spanning several administrative domains, a two-stage application of CryptoPan was developed that ensures that two different IPs pseudonymized by two different organizations is never mapped on the same IP.

Several transport mechanisms for sending data from a partner to the CarmentiS central can be used. An obvious is the Prelude framework [10] which offers strong authentication with X.509 certificates, integrity and confidentiality va

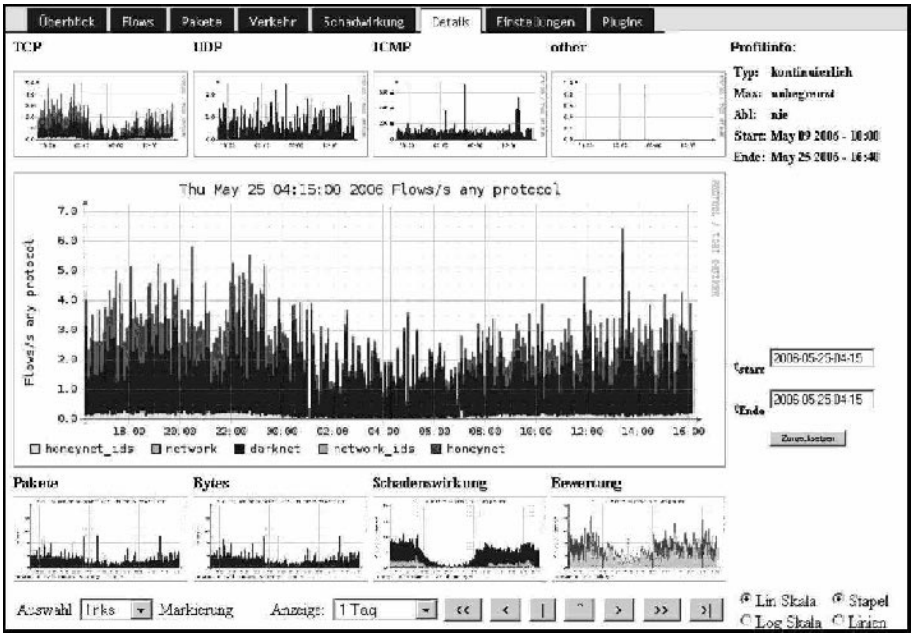

Fig. 3. Screenshot from CarmentiS Analysis Details Page

SSL encryption, and avoids data loss in case of network problems by output-buffering and auto-reconnect. Originally, however, Prelude only treats IDMEF data; plugins for the other data kinds used by CarmentiS have been developed.

3.2 Data storage and analysis

Due to the huge amount of data to be processed continuously and the requirements regarding processing time, the exclusive use of a relational database had to be ruled out from the beginning. Instead, a file-based approach that has proved feasible in productive use by SWITCH-CERT has been chosen. In this approach, which is geared towards Netflow data, sensor data is maintained in files representing five minute slices. After new sensor data are available, these are normalized and stored in an internal data format. In the following step the data is aggregated and supplied to the database. Active analysis methods are triggered and start processing the data. In particular, all active profiles defined by the CarmentiS analysts (see Section 2.2) are executed and the analysis results are stored. This carries a twofold benefit

– Analysts checking the result of standard profiles in regular intervals can access analysis results without time delay, because analysis is triggered automatically for new data. Thus, the analyst only has to wait for the processing of modified or newly created profiles.

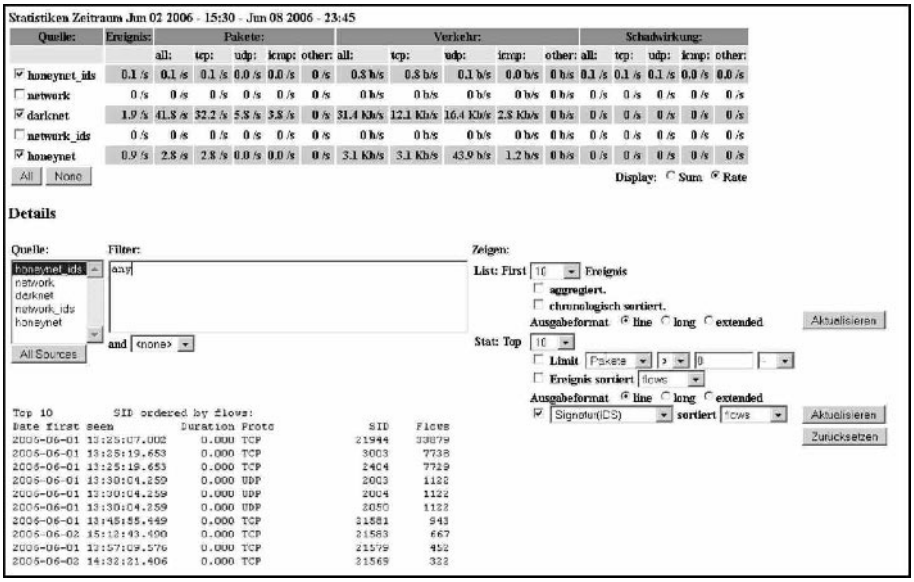

Fig. 4. Screenshot from CarmentiS Analysis Details Page (2)

– After original sensor data has to be discarded to create space for new data, analysis results for profiles executed while the data was available can still be accessed. Analysis results can be stored for a long time because they are very compact in comparison to original sensor data.

Because the NFSEN tools were developed exclusively for the processing and presentation of Netflow data, extensive modifications had to be made to be able to process IDS data and meta information.

3.3 Analysts Workbench

The analysts workbench was realized as a web front-end, programmed in 'Perl' and 'PHP'. The workbench offers the analysts different views and includes a graphical user interface for the tools for data processing.

Figure 3 shows the details tab for a detailed analysis of sensor data. The page is divided into two parts: The upper part gives a detailed view and allows to navigate through the sensor data. The content of the main graph can be selected by clicking into any of the smaller graphs which are arranged above and below the main graph (for instance in Figure 3 the selection is: *flows / any protocol*). The time window can be selected by a marking of the area of interest in the main graph with a pointing device. The pages are automatically refreshed every 5 minutes to update the graphs.

Figure 4 shows the lower part of the page, which contains all the necessary controls to process the sensor data within the given time window. For instance the top 10 IDS signatures during the given time window.

4 Outlook

With the basic CarmentiS infrastructure nearing completion, focus is shifting from planning and developing to using the infrastructure. Many aspects of (1) correlating data of different kinds and/or from different sources and (2) co-operatively analyzing data can only be researched with actual trials. Hence, the next step of the project will consist of conducting experiments to research these aspects, which are described in Section 4.1. Further work is also required to go from situation awareness achieved by CERTs taking part in the co-operative analysis towards achieving situation awareness for possible users of CarmentiS as identified in Section 2.1 and, finally, effective early warning; first steps in that direction are described in Section 4.2.

4.1 Experimenting with correlation and co-operative analysis

The following aspects of correlation and co-operative analysis are to be researched by conducting experiments with the basic CarmentiS infrastructure described in Section 3.

The practicalities of co-operative analysis To the best of our knowledge, CarmentiS is the first project to attempt co-operative analysis of sensor data between several CERT teams using a web-accessible analysts' workbench. So far, the part of the analysts' workbench that supports analysis of the collected data using an extended version of the NFSEN tools has been implemented. Only experiments in co-operative analysis will be able to tell how such analysis can be organized and which additional features of the analysts' workbench (see Section 3.3) must be added to support such co-operative analysis.

The role of pseudonymization As described in Section 2.2, capabilities for pseudonymization of IP numbers within sensor data have been tightly integrated into the CarmentiS infrastructure. Organization can thus supply sensor data which they might not be able to supply without pseudonymization because of privacy concerns. On the other hand, pseudonymization carried out on the level of organizations rather than the central data repository makes correlation harder or even impossible. The CarmentiS project will examine the effects of pseudonymization on correlation and co-operative analysis. At the same time, best-practice sharing between participating organizations regarding the legal aspects of submitting sensor data will provide invaluable input for creating the organizational framework of CarmentiS.

How to use sensor meta data As stated above, sensor data can only be interpreted sensibly in conjunction with further information about the sensor such as the export policy that was employed, sensor placement, configuration, etc. Such information can be submitted along with the actual sensor data within the common format defined within the CarmentiS project. Experiments will help the CarmentiS project in establishing the best methods to integrate such meta data into analysis.

4.2 From co-operative analysis to early warning

CarmentiS partners taking part in the co-operative analysis will be the first to benefit from CarmentiS: at the very least, they will have an additional tool useful for incident handling: events observed within the own network can be compared with events seen on other networks. The next step must be to support situation awareness and early warning also for possible users of the CarmentiS system not involved in cooperative analysis.

Creating user-specific situation awareness In order to support the CarmentiS stakeholders with information about the current IT-security situation, a web-portal will be implemented. At the moment, we envision to offer different views for each group of stakeholders, containing statistical information, reports and warnings prepared by the analysts, information provided by partners and the CERT community as well as publicly available information sources. How different users can be served best, however, is still subject to research.

Improving early warning with automated analysis techniques As described in Section 2.2, automated analysis techniques are to be used to notify analysts of events of interest that warrant further examination. Keeping in line with the modular approach that characterizes the CarmentiS approach so far, a well-defined interface for integrating automated analysis techniques into the CarmentiS framework will be implemented.

Preliminary results have shown, that simple threshold schemes are quiet efficient compared to scientific more elaborate methods i.e. neural networks or statistical analysis [11], but that the later can identify more complex behaviours simple methods cannot. Therefore we plan to start with the implementation of two algorithms:

1. A threshold scheme based on volume classes high, medium, low. Instead of concentrating on the raw data records we will focus on the number of raw data record sets that conform to pattern created by association rule mining [12] in all records. To achieve usability, the past will be used to determine the appropriate volume class for such pattern. The creation of new pattern will, by itself, trigger manual analysis to assess the relevance of that pattern.
2. The second approach will probably concentrate on hidden markov models [13] based on further evaluation of available scientific results.

5 Conclusion

CarmentiS provides a co-operative approach towards situation awareness and early warning in the Internet. At its core is an infrastructure and organizational framework for sharing, correlating and cooperatively analyzing sensor data. Development and deployment of the basic infrastructure components have progressed such that experiments in collecting sensor data from several institutions, correlating and co-operatively analyzing this data can commence. Experiences collected with these experiments are crucial for the definition of a viable organziational framework and the further development of the exisiting infrastructure.

Thanks to its unique co-operative approach, CarmentiS has unmatched opportunities in pooling both sensor data as well as resources and expertise for analyzing that data. It has thus the potential to serve both as a platform for conducting state-of-the-art research regarding early warning and as viable basis for a national early warning and information system.

References

[1] DShield: DShield Website, (`http://www.dshield.org/`)

[2] MyNetWatchman: MyNetWatchman Website, (`http://www.mynetwatchman.com/`)

[3] eCSIRT.net: eCSIRT.net Website, (`http://ecsirt.net/`)

[4] IMS: Internet Motion Sensor (IMS) Website, (`http://ims.eecs.umich.edu/index.html`)

[5] BSI: German Federal Office of Information Security (Website), (`http://www.bsi.de/`)

[6] Danyliw, R., Meijer, J., Demchenko, J.: The Incident Object Description Exchange Format Data Model and XML Implementation. Technical report, IETF Incident Handling WG, `http://www.cert.org/ietf/inch/docs/draft-ietf-inch-iodef-06.txt` (2006)

[7] Bundesamt für Sicherheit in der Informationstechnologie (BSI) `http://www.cert-verbund.de/sirios/index.html` (in German): Cert Verbund - SIRIOS. (2005)

[8] Debar, H., Curry, D., Feinstein, B.: The Intrusion Detection Message Exchange Format. Technical report, IETF Intrusion Detection Exchange Format Working Group, `http://www.ietf.org/internet-drafts/draft-ietf-idwg-idmef-xml-16.txt` (2006)

[9] Cryptopan: Crypto-PAn : Cryptography-based Prefix-preserving Anonymization, (`http://www-static.cc.gatech.edu/computing/Telecomm/cryptopan/`)

[10] Prelude: Prelude Website, (`http://www.prelude-ids.org/`)

[11] Chyssler, T., Nadjm-Tehrani, S., Burschka, S., Burbeck, K.: Reduction and Correlation in Defence of IP Networks. In: Proceedings of the 13th IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). (2004) 229–234

[12] Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules Between Sets of Items in Large Databases. In: Proceedings of the ACM SIGMOD Conference on Management of Data. (1993) 207–216

[13] Jha, S., Tan, K., Maxion, R.: Markov Chains, Classifiers, and Intrusion Detection. In: 14th IEEE Computer Security Foundations Workshop (CSFW). (2001)