Technology Roadmap and Dynamic Protection for Big Data Application in the Healthcare Domain

Gangappa B Demannavar
M.Tech 2nd semester
Department of CSE
City Engineering College

Girish G A
Associate professor
Department of CSE
City Engineering College

Abstract—**Big Data technologies can be used to improve the quality and efficiency of healthcare delivery. The highest impact of Big Data applications is expected when data from various healthcare areas, such as clinical, administrative, financial, or outcome data, can be integrated.. For enabling the seamless access several technical requirements, such as data digitalization, semantic annotation, data sharing, data privacy and security as well as data quality need to be addressed. Critical Care IT systems such as life supportdevices, vitals monitoring systems, information systems that provide point of care guidance to care teams are a key component of a lifesaving effort in Healthcare, has created new challenges and the point in time detection methods at the hospitals are no longer effective and pose a big threat to the critical care systems. To maintain the availability and integrity of these critical care systems, new adaptive, learning security defense systems are required that not only learns from the traffic entering the hospital, but also proactively learns from the traffic worldwide. Cisco's Cloud web security (CWS) provides industry-leading security and control for the distributed enterprise by protecting users everywhere, anytime through Cisco worldwide threat intelligence, advanced threat defense capabilities, and roaming user protection. In this paper, we introduce a detailed analysis of these technical requirements and show how the results of our analysis lead towards a technical roadmap for Big Data in the healthcare domain and how big Data Analytics is used in combination with other security capabilities to proactively identify threats and prevent wide spread damage to healthcare critical assets.**

Keywords—Big Data; technical requirements; data digitalization; semantic annotation; data integration; data privacy and security; data quality Healthcare, Security, Critical Care, Big Data Analytics, Behavior Analysis, Machine Learning, Malware,

I.INTRODUCTION

The healthcare domain faces tremendous productivity challenges. Due to the changing patient demographics as well as the increasing healthcare costs, there is a clear need for cost efficiency, improved quality of care, and broader healthcare services.Recent studies [1-4] highlight that Big Data technologies and health data analytics are being used to address the efficiency and quality challenges in the healthcare domain. For instance, by aggregating and analyzing health data from disparate sources, such as clinical, financial and administrative data, the outcome of treatments in relation to the resource utilization can be monitored. This aggregation in turn helps to improve the efficiency of care. Moreover, the identification of high-risk patients and predictive models leading towards proactive patient care allows to improve the quality of care.

After performing a comprehensive analysis of domain needs and requirements, we found that the highest impact of Big Data applications in the healthcare domain is achievable when it becomes possible to not only acquire data from one single but various data sources such that different aspects from the various sectors can be combined to gain new insights,Therefore, the availability and integration of all related health data sources, such as clinical data, claims, cost and administrative data, pharmaceutical and R&D data, patient behavior and sentiment data as well as the health data on the web, is of high relevance [5].However, as of today, the access to health data is only possible in a very constrained manner. In order to enable seamless access to healthcare data, several technical requirements need to be addressed such as: 1) health data is documented in digitalized manner without imposing extra-effort for physicians 2) the content of unstructured health data (such as images or reports) is enhanced by semantic annotation 3) data silos are conquered by means of efficient technologies for semantic data storage and exchange 4) technical means backed by legal frameworks ensure the regulated sharing and exchange of health data, and 5) means for assessing and improving the data quality are available.Today, in healthcare, IT systems plays a central role in critical clinical care. Some of their roles include collecting vital information, augmenting human life support, enabling communication, archiving and information sharing. Any attempt to disrupt the availability and integrity of these systems has far reaching consequences in healthcare. there are new challenges and threats that the critical care IT systems must be aware of.The main aim of this paper is to describe and analyze these technical requirements in detail as well as indicate the state-of-the-art and the power of big data to perform behavioral analysis, anomaly detection, evasion resistance, rapid Detection services using flow based, signature based, behavior based and full packet capture models to identify threats.

II. METHODOLOGY

For developing the technology roadmap for Big Data applications in the healthcare domain, we followed a process of several steps:

1. In the first step, we analyzed all the needs and requirements in the healthcare domain that could be addressed by means of healthcare analytics and Big Data technology. For that, we accomplished a *review ofavailable literature, internet sources and market studies* that guided us in developing a comprehensive questionnaire.
2. The questionnaire was used to conduct *12 semi-structuredinterviews* with representatives of all stakeholder groupsof the healthcare domain, such as patients, clinicians, hospital operators, pharmaceutical companies, R&D, payors, and medical product providers. The questionnaire focused on three aspects: a) we asked the interviewees about user needs that could be addressed by means of healthcare IT, b) we asked the interviewees to evaluate a list of precompiled Big Data application scenarios (which we found within our review) as well as to describe other promising Big Data scenarios they are aware of and c) we reviewed with them a list of possible constraints that are hindering the successful implementation of Big Data scenarios in the healthcare domain.
3. By *clustering and ranking the discussed use casescenarios*, we derived a set of six high-level applicationscenarios. By analyzing these application scenarios, we identified relevant constraints and requirements that need to be in place for the successful implementation of the scenarios. By aligning this initial list of constraints and requirements with the input from our interviews, a final list of constraints/requirements was compiled.
4. In our further analysis, we distinguished between technical and business-related requirements[3]. The technical requirements (which we also call *enablingtechnologies*) were analyzed in further detail bydescribing the AS-IS and TO-BE[4] situation, by investigating the required functionalities, the available technologies as well as by identifying open R&D questions. When needed, we conducted further expert interviews. The summary of this analysis is described in the following two sections[5] III and IV.
5. Besides our analysis of enabling technologies, the investigation of *future opportunities* associated with Big Dataapplications in the healthcare domain is needed. This
6. The final step within the roadmap development is the temporal alignment and ranking of technical requirements described in Steps 4 and 5. As the adoption of new technologies depends on the degree to which the identified business requirements can be addressed, we determine how the business requirements can be influenced and by whom they need to be addressed.

Step 5 and 6 is part of our future work.

III.ENABLING TECHNOLOG IES

In order to establish the basis for wide-spread usage of Big Data applications in the healthcare domain, several technical requirements need to be addressed. We labeled these technologies as “enabling”, because they establish the technical foundation for subsequent Big Data applications. In this way, *enabling technologies* cover all health-specific data management technologies that ensure that the various heterogeneous health data pools can be easily accessed and that the health data is integrated and available. Within our analysis of technical requirements, we distinguish enabling technologies from *value creating technologies* that are needed to elaborate concrete Big Data-based business opportunities. The differentiation between enabling and value-creating technologies is needed for our further analysis in order to indicate how and to which extent the various technologies depend on each other: Enabling technologies often require long-term investments from various partners without immediate business potential, while value-creating technologies relate to concrete business opportunities that assume that the various data sources are available (by means of enabling technologies).

A. Data Digitalization

Some years ago, data digitalization was a huge problem, but today it is progressively becoming a less important problem from a technical point of view. Nevertheless digital data is still not available everywhere. The extent to which healthcare Information Technology (IT) systems are in use differs across and also within countries. A study showed that on average 55% of healthcare providers in Germany use healthcare IT within primary care settings and 60% in secondary care settings [9]. For the United Kingdom, the numbers are different: 63% of providers rely on healthcare IT in primary and only 15% in secondary healthcare [9].

In a perfect scenario, data collected for primary care settings should be available in an annotated, curated and high quality manner for secondary use.

B. Semantic Annotation

When working with health-related data in general and Big health-related Data in particular, one is facing the challenge

[3] Three business-related requirements were identified: a) lack of promising business cases, b) the need for high investments and c) need for value-based incentive system.

[4] TO-BE Situation refers to the year 2020

[5] The complete analysis can be accessed under http://www.big-project.eu/is done by analyzing the technology required for implementing the selected high-level scenarios.

of data heterogeneity (reports, lab reports, images, sensor data, etc.). In addition, large amounts of information are captured in unstructured formats (e.g. reports or images). The International Data Corporation (IDC) market research institute estimates that in the upcoming years 90% of health data willbe provided in unstructured formats [6]. Semantic annotation is described as a possible solution for processing heterogeneous unstructured data seamlessly.Besides a small set of standardized metadata, such as the DICOM header, that provides the basis for the exchange and management of documents, the content of unstructured information is in general not provided in standardized formats. For example, reporting in radiology is still conducted as free text. The reading and interpretation of such data is accomplished manually by individual clinicians. Without semantic annotations, it is not possible to process the content of unstructured data automatically. Therefore, a holistic analysis of the patient's status is hindered.The envisioned impact of semantic annotation is very promising. In order to automatically align related data sets, their content needs to be represented explicitly and consistently. This means, semantic annotation relies on commonly used vocabularies or ontologies.

C. Data Sharing

As of today, a lot of health data is stored in data silos. A seamless exchange and aggregation of the data often relies on individualized solutions due to the lack of standards and flexible interfaces as well as the heterogeneous nature of the data. In comparison to the degree of healthcare IT adoption, the incorporation of seamless information exchange is far less advanced [9]. On average, for instance, in Germany less than 23%, in the UK less than 46% and in the US less than 36% of the healthcare providers use healthcare information exchange technology [9].

In terms of analyzing the current state of health data sharing, several deficiencies have to be emphasized: First, the data exchange within one healthcare provider is complicated due to the usage of different information systems in different departments. Although it is feasible from a technical point of view to exchange health data by e.g. using HL7[6] CDA (Health Level 7 – Clinical Document Architecture), as of today health data is hardly shared across organizations due to non-technical reasons. Moreover, the healthcare domain lacks internationally accepted coding systems. Even the ICD[7] (International Classification of Diseases), which is broadly accepted, is used in country-specific adaptations only. Other promising systems still lack acceptance (e.g. SNOMED Clinical Terms[8]). In terms of underlying means for data representation, existing Electronic Health Record (EHR) systems mainly provide a case-centric instead of a patient-centric view, which hinders longitudinal health data integration.

To enable seamless health data exchange, standardized data models for clinical data as well as coding schemes for labeling content need to be agreed upon.

D. Data Privacy and Security

As health data is private data, its processing needs to follow high data security and privacy constraints. Therefore,there is a strong need for technical infrastructures, legal processes for data sharing and communication as well as for the implementation of suitable organizational processes that enable secure and transparent health data sharing.Since the importance of data and especially of Big Data is increasing and becoming a competitive company value, data security, risk management and data privacy requirements are becoming more and more important. For ensuring transparent and secure data sharing, challenges on four different levels, namely the legal, technical, organizational and social level, have to be taken into account: 1) In the European Union the huge importance of (health) data privacy and security is underlined by Directive 95/46/EC [7]. Based on this minimal *legal* standard, all member states were called to implement anational data protection law. This is one of the major issues concerning data privacy and security within the European Union. There does not exist one common legal framework for all member states. Each member state issued its own data protection law. 2) From a *technical point of view*, the implementation of secure Big Data applications with respect to data privacy is supposed to be possible with the available technologies (e.g. Integrating the Healthcare Enterprise (IHE) initiative[9]). Other privacy enhancing techniques like anonymization or pseudonymization of personal health-related data are shown and described in [8]. 3) From an *organizational point of view*, the storage, processing, access,and protection of Big health Data has to be regulated on different levels (e.g. institutional, regional, national, international). 4) From a *social point of view*, the huge benefits of Big health Data need to be communicated and promoted.

When talking about anonymization in Big health-related Data context, one always has to keep in mind the possibility of unintentional re-identification of individuals, when aggregating anonymized Big health Data from different sources. This problem results from the additional knowledge gained by the merging of different data sources.In order to resolve the mentioned issues, a common legal framework within the European Union regulating the intramural and extramural exchange and processing of health-related (Big) Data is needed.

E. Data Quality

The data quality in Big health Data application depends on the quality of data of the original data sources. Usually medical product providers are not responsible for the quality (e.g completeness, accuracy) of the data collected and documented in hospitals. Nevertheless, they provide tools for data collection, documentation and analysis. These tools can help to support hospitals and healthcare providers to ensure that data is complete. Additionally, the tools can help to improve data quality (e.g. plausibility checks, mandatory items). A major issue regarding data quality, which directly links it to data digitalization, are media disruptions. The more media disruptions exist in a process chain, the worse the data quality gets.

Poor data quality is a major limitation to the benefit of (Big) Data analysis and the quality of healthcare itself.

Therefore, before analyzing data two assessment steps have to be carried out: a) the quality of the data sets used for Big Data applications has to be evaluated and b) quality has to be improved by using appropriate approaches (e.g. multiple data imputation).

IV. TECHNICAL ROADMAP DESCRIPTION

In this section, we will analyze the required technologies for addressing the above mentioned challenges. However, one needs to keep in mind that the availability of technology is not sufficient to solve the mentioned challenges, but dedicated processes, standards or frameworks need to be available and implemented. As these non-technical aspects might influence the temporal axis of our technical roadmap, we decided to integrate those aspects within our analysis, but to indicate them explicitly. Figure 1 provides a consolidated view of the main enabling and value-creating technologies, which are described in the following sections. These technologies may be already available or still open R&D topics. Enabling technologies are referred to as technologies, which establish the technical foundation for subsequent Big Data applications, while value creating technologies create a concrete business value. For a better understanding, we highlighted (italic) the catchwords also appearing in Figure 1 throughout the text and tagged them (e.g. *A1*) in order to find them quicker in the figure.

A. Data Digitalization

Digitization of medical data has a huge impact on every aspect of the healthcare domain: healthcare delivery, health management, healthcare policy making, and administration. Digital health-related data supports integrated healthcare delivery, highly automated data processing, research, as well as medical routine. Various technologies for the digital capturing of health data, such as Hospital Information System (HIS), Radiology Information System (RIS), Picture Archiving and Communication System (PACS), etc. are already in place.In order to improve the health data documentation process by reducing the extra effort for physicians, several technologies in the domain of m*edical speech recognition(A1)* systems provide means for implementing non-paperbased documentation.

B. Semantic Annotation

To enable semantic annotation for medical data, some major requirements have to be fulfilled. Annotation requirements should be based on medical context understanding. Therefore, a fully covered medical terminology with an attached semantic classification is needed as well as *medical information extraction (B1)* techniques for themedical domain, which allow to integrate external domain knowledge. Also an annotation process, which makes the semantic of the health data explicit, is required.

As of today, there are some technologies available, which meet the requirements mentioned above. Nevertheless they are not yet widely used and/or accepted on a broader basis. *Semantic medical terminologies (B2)* (e.g. Unified MedicalLanguage System (UMLS), SNOMED Clinical Terms (SNOMED CT)), which contain a lot of medical knowledge and information have several shortcomings (e.g. lack of multilingual concepts [10]). *Semi-structured reportingstandards (B3)* (e.g. RSNA reporting standards for Radiology)are not commonly used.

There are still some (partially) open Research and Development (R&D) questions concerning semantic annotation. First *standardized medical annotation frameworks(B4)* and *enriched semantic medical terminologies (B5)* arerequired. Available terminologies have problems and need to be extended regarding, quality and/or quantity and semantic description of concepts. Also *semi-structured reportingalgorithms (B6)* and *image-understanding algorithms (B7)* should be targeted by research.

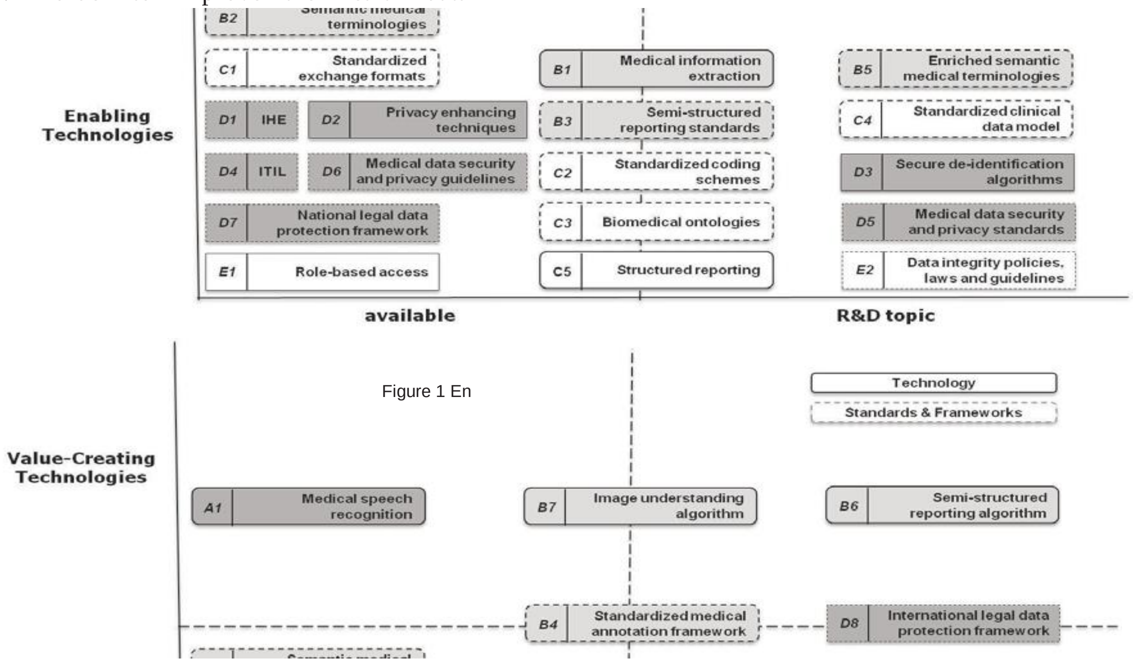
B2 Semantic medical terminologies
C1 Standardized exchange formats
B1 Medical information extraction
B5 Enriched semantic medical terminologies
Enabling Technologies
D1 IHE
D2 Privacy enhancing techniques
B3 Semi-structured reporting standards
C4 Standardized clinical data model
D4 ITIL
D6 Medical data security and privacy guidelines
C2 Standardized coding schemes
D3 Secure de-identification algorithms
D7 National legal data protection framework
C3 Biomedical ontologies
D5 Medical data security and privacy standards
E1 Role-based access
C5 Structured reporting
E2 Data integrity policies, laws and guidelines
available
R&D topic
Technology
Standards & Frameworks
Value-Creating Technologies
A1 Medical speech recognition
B7 Image understanding algorithm
B6 Semi-structured reporting algorithm
B4 Standardized medical annotation framework
D8 International legal data protection framework
Semantic medical

Figure 1 En

C. Data Sharing

To enable semantic data integration and data sharing, common standards for representing data are strongly required. Therefore, standardized data storage and exchange formats as well as a structured and standardized representation of health data are needed. Commonly accepted and used coding systems, a complete, patient-centric and longitudinal patient data representation as well as proper semantic annotation algorithms are required to facilitate and support seamless data sharing and exchange.

There are already technologies available, which meet the requirements mentioned above: Health Level 7 (HL7) provides commonly accepted *standardized exchange formats(C1)* (HL7v2.x and v3.0) as well as document standards (HL7Clinical Document Architecture (CDA)). To provide a common meaning and understanding of medical data, *standardized coding schemes (C2)* and terminologies arerequired. They allow a common representation of structured data (e.g. ICD or LOINC[10]). Coding systems for more granular health data like SNOMED CT are available even though with several open issues regarding consistency, performance and unambiguous representation [11-15]. *Biomedical ontologies(C3)* (e.g. Gene-Ontology[11]) as well as *standardized clinical data models (C4)* (e.g. HL7 Reference Information Model(RIM)[12] are already available.

A majority of the technologies mentioned above have several issues, which need to be targeted by further research: Most of the existing ontologies are available in English but lack in multilingual representations and, therefore, in usability. Regarding standardized data models we note that HL7 RIM has certain issues concerning the ontological consistency [16]. Research activities for the development of an integrated patient data model on the basis of well-defined ontologies are on-going (e.g. Model for Clinical Information [17]). There are numerous other initiatives developing data models for improved data sharing in the context of clinical studies. Prominent examples are the Translational Medicine Ontology [17] or models developed by the Clinical Data Interchange Standards Consortium[13] or within the EHR4CR project[14]. Another important target for research are tools facilitating *structured reporting (C5)* which do not put extra work forclinicians and can be seamlessly integrated in the clinical workflow.

D. Data Security and Privacy

The Cloud Security Alliance [18] listed major Big Data security and privacy challenges, which include secure computations in distributed programming frameworks, secure data storage and transaction logs, real-time security and compliance monitoring, scalable and composable privacy-preserving data managing and analysis approaches and granular access control and audits.

As of today, some technologies enhancing data privacy and security are already available. The Integrating the Healthcare Enterprise (*IHE (D1)*)[15] initiative enables plug-and-play and secure access to health information whenever needed. IHE also promotes the use of well-established and internationally accepted standards, such as DICOM or HL7. [19] lists some *privacy enhancing technologies (D2)* (e.g. Peterson Approach, Pommerening Approaches, Electronic Health Card). Some of these approaches use pseudonymization techniques, which is important especially for longitudinal medical research. Here it is sometimes necessary to re-identify study subjects in order to, for example, communicate important study results. As mentioned earlier, unintentional re-identification of individuals becomes a real problem, when integrating linked data from various sources. Therefore, the real dimension of anonymity has to be evaluated and *secure de-identification algorithms (D3),* such as the k-anonymity approach have to be used or developed. To support IT service management, *ITIL*[16]*(D4)* is a widely accepted approach that provides a cohesive set of best practices including guidelines for data security and privacy.

For enhancing data privacy and security as well as for the other requirements listed in this article, internationally accepted *medical data security and privacy standards (D5)* are needed. Therefore existing *medical data security and privacyguidelines (D6)* can serve as a point of reference. As there willalways be efforts for cracking the existing algorithms in order to collect personal health data, secure data de-identification approaches need to be continuously evolved. One major requirement for enhancing data privacy and security is a rigid and proper *national legal data protection framework (D7)* as well as *international legal data protection framework (D8)*.

E. Data Quality

According to [20], in order to improve data quality, the following requirements need to be met: a) a controlled, *role-based access (E1)* can hinder unauthorized reading andwriting actions b) data dictionaries have to be used to achieve a common terminology c) standardized formats ensure consistency and follow directly from the last point mentioned d) *data integrity policies, laws and guidelines (E2)* are needed and have to be followed.

V. HEALTHCARE CRITICAL IT SYSTEMS

Healthcare IT systems are an important element in modern day Clinical care. On a very high level, we can classify the healthcare IT systems into four categories.

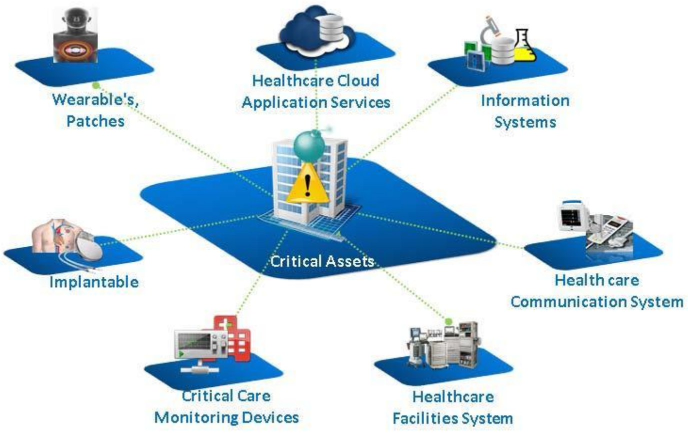
Wearable's, Patches
Healthcare Cloud Application Services
Information Systems
Critical Assets
Implantable
Health care Communication System
Critical Care Monitoring Devices
Healthcare Facilities System

Figure 2. Examples of Critical Care IT systems

A. Human Support Systems

Human Support Systems are life support systems that are critical for a patient's continued quality of life. They complement the human organ functions. An example of such system is an insulin pump or a pacemaker.hospital IT systems and increases the attack surface significantly. Unpatched systems, multi user, interaction with social and other applications can be a source of enhancement vulnerabilities.

B. Vitals Collection/monitoring Systems

Effective treatment decisions rely on available data and hence collection of the vital information of patients is very important in clinical care. Examples of a vitals collection/monitoring systems include a pulse oximetry, Electro cardiogram (ECG).

C. Information Systems

Clinical care involves interaction with multiple care team members, who need access to the same information about the patient. Information systems are used to store, share anupdate such information. Examples of such systems include Electronic medical records (EMR), PACS (picture archiving and communication system)

D. Facilities and Support Systems

Maintaining optimal environment is very important in clinical care to ensure the quality of products used in care. For example, blood supplies or medications have temperature requirements. The Facilities and Support Systems enable meeting these requirements. Other examples include communication systems, room systems such as hospital beds and control systems

VI. SECURITY CHALLENGESAND IMPACT

With the widespread adoption of mobile, social and cloud on one hand and the sophistication of the malwares on the other hand, security threats are prevalent in most industries today. While the availability and integrity of information is important in all industries, it has farther reaching implications in healthcare.

The landscape of threats is influenced by multiple factors, but we will focus only on user preference and usage changes and advances in cybercrime in this paper. Some of them include:

A. User preference and Usage Models

As users of healthcare start to use a combination of the megatrends of mobile, social and cloud, it is much easier for hackers to get access to systems to infiltrate the system. The boundaries of the healthcare systems are constantly evolving as patients have access to their information from the home PCs and mobile devices. The control on what type of devices that system is accessed is no more in the control of thehospital IT systems and increases the attack surface significantly. Unpatched systems, multi user, interaction with social and other applications can be a source of enhancement vulnerabilities.

B. Medical Device and IT systems Innovation

The medical industry finds new innovations for finding cure, enabling quality of life and improves efficiencies. Today, medical devices range from devices that are implanted in a patient to devices components that are on the cloud.

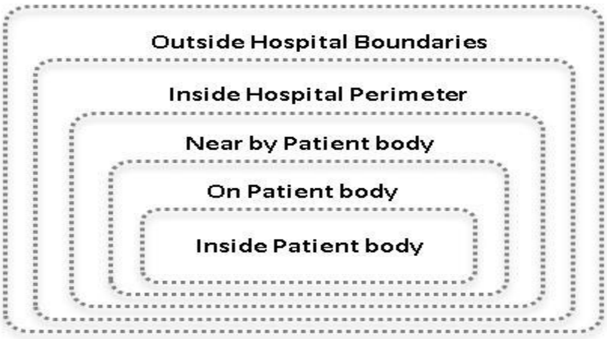

Figure 3. Block model of system security proximity of critical care assets

The sensors inside the body have very limited resources in terms of power, storage and processing [2] and hence have limited capabilities to self-protect. As the devices are closer to the body proximity, the impact can be more dangerous. Such possibilities of attacks on medical devices such as insulin pumps [3, 4] have been showed off by hackers such as Barnaby Jack and Jay Radcliffe in the past.

C. Sophistication of Malwares

The threat landscape has completely evolved over time and today, the sophistication of malware has posed significant questions on traditional approaches of endpoint security using Antivirus (AV) software and perimeter defenses. The attack models have changed from simple attacks on PCs to focused large scale attacks on targeted entities using approaches such as techniques called Advanced Persistent Threats. Such attacks leverage Zero data exploits, Spear Phishing, Watering hole models, encrypted side channel methods are used to infect systems.

D. AV software Evasion Techniques

The malwares, in addition to attacking systems, also has evolved in the use of Evasion Techniques such as malware packing, obfuscation, and polymorphism. These techniquesevade from the signature based antivirus software protection and increase the risk of attack on critical systems.

E. Leveraging Malware Models with sensitive situations

New forms of malware such as Ransomware when tied with a healthcare scenario can become a powerful attack model. Let’s look at the use of ransomware in a Healthcare situation. Ransomware [5] is a type of malware which restricts access to the computer system that it affects, and demands a ransom paid to the creators of the malware in order for the restriction to be removed. Assume a critical care IT system (a device or an information system that is controlling the care) is infected with a ransomware malware; the need to respond to malware creator might be higher due to the life threatening situation as compared to a normal user who has a home PC that has some level of data that is important, but is not critical.

F. New Malware Service Models

New malware service models are emerging where malware is provided as a service model and thereby enable many entities to fine tune and provide variations of malware quickly. This makes signature detection extremely challenging.

VII. THE AUGMENTED MULTI LAYERED SECURITY MODEL

At each critical element, a threat model needs to be defined and a corresponding protection model must be defined. The protection model is a function of various protection shields that the asset has to protect itself from attacks from known malwares and new malwares.

Protection Shield for a critical asset = fn (Self-protection capabilities, Endpoint Protection capabilities, Perimeter defense capabilities, Global intelligence)

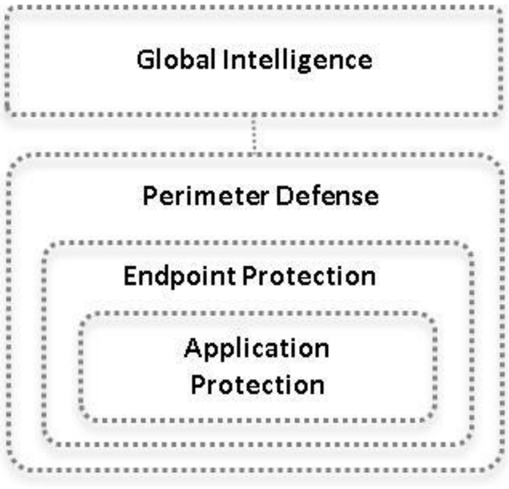

Figure 4. Protection shield model for critical asset

The self-protection capabilities include encryption, secure channels, secure data access and storage. The endpoint protection capabilities include antivirus software, secure access to device, containerization, device firewall. Perimeter Defense include firewall, intrusion detection and prevention systems.

The global intelligence layer provides the intelligence based on the learning from worldwide traffic and learning from anomalies, behavior analysis, and deep packet analysis. This model augments the protection mechanism and protects the critical assets using a multi vector/layer approach and makes up for the vulnerability window if new malware sneak in bypassing the perimeter defense.

VIIITHE NEED FOR DYNAMIC PROTECTION

As more care team members and patients leverage mobile, social and access content from the various web sites, smarter malware that uses obfuscation and packing bypasses the perimeter defenses and enters the enterprise. Leveraging zero day exploits and other attack techniques, they go undetected by traditional signature based detection software.

In addition, with new malware as a service models and polymorphic behaviors makes the malwares extremely dynamic.

To tackle such scenarios, a dynamic model is required. Netflow/HTTP anomaly detection, protocol metadata forensics, protocol anomaly detection, full packet forensics, behavior analysis, sand boxing are few techniques used in this dynamic model. It also uses advanced statistical modeling and machine learning to make more accurate determinations, and responds to new threats as they emerge.

IX. CISCO WEB SECURITY AND CISCO MANAGEDTHREAT DEFENSE

Cisco Cloud Web Security (CWS) [6] provides industry-leading security and control for the distributed enterprise. Through a combination of best-in-class uptime, unmatched zero-day threat protection, advanced malware protection, and cutting-edge analytics, Cisco CWS provides continuous monitoring and analysis across the extended network and throughout the full attack continuum: before, during, and after an attack.

CWS performs active, continuous monitoring for threats that have penetrated defenses, accurate and fast identification of threats, stops the spread of an attack, consistent and reliable spotting of new exploits by focusing on anomalous behavior. It makes use of advanced statistical modeling and machine learning that make more accurate determinations.

Leveraging Cisco’s Security Intelligence Operations (SIO), [7] it brings together global security intelligence from the cloud with local intelligence on a customer premise toprotect devices and information systems against advanced cyber threats. Cisco Security Intelligence Operations (SIO) provides a 24-hour view into global traffic activity that enables Cisco to analyze anomalies, uncover new threats, and monitor traffic trends. Cisco SIO generates new rules and updates every three to five minutes, providing threat defense hours and even days ahead of competitors.

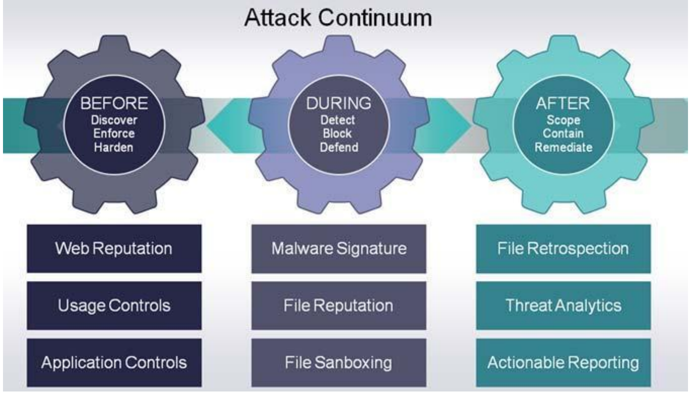

Figure 5. The Attack continuum

Big data analytics is used to generate real-time threat intelligence. The system analyses daily more than 100 TB of security intelligence, and 16 billion web requests to detect and mitigate threats. [8]

It also has granular visibility and control of more than 150,000 applications and micro-applications. It defends against zero-day web malware through dynamic reputation and real-time threat intelligence from Cisco SIO. All inbound web traffic to the healthcare entity is scanned in real time using context-aware scanning engines to identify and block untrusted domains.

CWS identifies unknown, unusual behaviors through Cisco Outbreak Intelligence, a heuristics-based engine that runs webpage components in a highly secure environment before permitting user access.

Cisco Cognitive Threat Analytics [9] is a cloud-based solution that reduces time to discovery of threats operating inside the network. It addresses gaps in perimeter-based defenses by identifying the symptoms of a malware infection or data breach using behavioral analysis and anomaly detection.

Unlike traditional monitoring systems, Cisco Cognitive Threat Analytics relies on advanced statistical modeling and machine learning to independently identify new threats, learn from what it sees, and adapt over time.

Anomalous traffic patterns and suspected incidents are escalated to a trained Cisco security investigator in one of the global security operations center for further analysis.Administrators can select specific categories for intelligent HTTPS inspection, and a single management interface delivers global control and comprehensive reporting. When using Cisco CWS, users are protected everywhere, all the time, through Cisco's worldwide threat intelligence footprint. As a cloud service, Cisco CWS offers ease of deployment, and the ability to centrally set and enforce policies for an entire organization, regardless of where users are located. Cisco CWS also uses the power of cloud computing to stop threats.

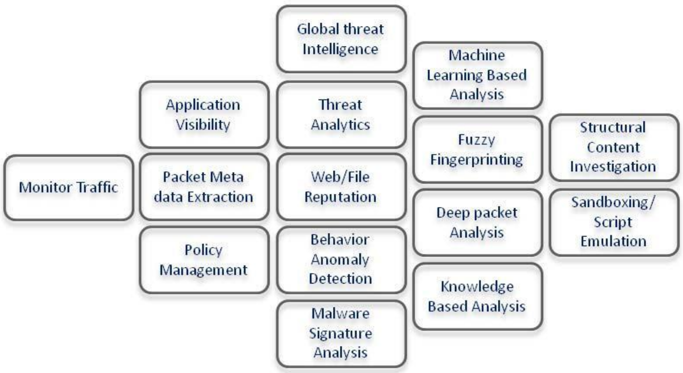
Global threat Intelligence
Machine Learning Based Analysis
Application Visibility
Threat Analytics
Fuzzy Fingerprinting
Structural Content Investigation
Monitor Traffic
Packet Meta data Extraction
Web/File Reputation
Deep packet Analysis
Sandboxing/ Script Emulation
Policy Management
Behavior Anomaly Detection
Knowledge Based Analysis
Malware Signature Analysis

Figure 6. The Analysis techniques to identify malware threats

Leveraging solutions such as CWS that use context aware scanning, Machine learning algorithms and predictive analytics to detect possible threats in real-time can help protect critical care assets before any harm is caused.

X. CONCLUSION

Our study has shown that the impact of Big Data applications is very promising, but relies on the availability of healthcare data. Thus, several technical challenges, such as data digitalization, semantic annotation, data sharing, data quality anddata privacy and security, need to be investigated. Within our study, we identified the state-of-the-art

Protecting critical care asset in Healthcare IT is central to delivering care and enabling quality of life for patients. The dynamic protection model leveraging big data analytics is critical to get ahead of the ever evolving malware threat landscape. Techniques such as deep content Analysis, Structural content investigation and virtualized script emulation and leveraging the machine learned knowledge can help identify threats early in the cycle before it attacks critical care assets.

REFERENCES

[1] Frost and Sullivan, U.S. Hospital Health Data Analytics Market, 2012.

[2] McKinsey and Company. Big data: The next frontier for innovation, competition, and productivity. 2011.

[3] P. Groves, B. Kayyali, D. Knott and S. Van Kuiken. The 'big data' revolution in healthcare. McKinsey & Company. 2013

[4] M. Porter and E.OlmsteadTeisberg. Redefining Health Care: Creating Value-Based Competition on Results. Boston: Harvard Business Review Press, 2006.

[5] S. Zillner et al. *D2.3.1 First Draft of Sector's Requisites*. Public Deliverable of the EU-Project BIG (318062; ICT-2011.4.4), 2013.

[6] LünendonkGmbh. Trendpapier 2013: Big Data beiKrankenversicherungen. Bewältigung der Datenmengen in einemverändertenGesundheitswesen. online available, 2013.

[7] European Parliament and the Council of the European Union. Directive 95/46/EC of the European Parliament and of the Council 1995, [http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML], 1995.

[8] International Organization for Standardization ISO/TS 25237:2008 Health informatics - Pseudonymization. 1.edition, Geneva, 2008

[9] Accenture. Connected Health: The Drive to Integrated Healthcare Delivery. Online: www.acccenture.com/connectedhealthstudy, 2012.

[10] J. Ingenerf. Die Referenzterminologie SNOMED CT_: von theoretischenBetrachtungenbiszurpraktischenImplementierung. Neu-Isenburg, Germany, 2007.