Technical Challenges of Syndromic Surveillance System Deployment in a Health Information Exchange

Julio C. Silva[1], Charles Cox[2], Dino P. Rumoro[1], Shital C. Shah[1], Marilyn M. Hallock[1], Gillian S. Gibbs*[1], Gordon M. Trenholme[1] and Michael J. Waddell[3]

[1]Rush University, Chicago, IL, USA; [2]Metropolitan Chicago Healthcare Council, Chicago, IL, USA; [3]Pangaea Information Technologies, Ltd., Chicago, IL, USA

Objective

The objective is to describe the technical process, challenges, and lessons learned in scaling up from a local to regional syndromic surveillance system using the MetroChicago Health Information Exchange (HIE) and Geographic Utilization of Artificial Intelligence in Real-Time for Disease Identification and Alert Notification (GUARDIAN) collaborative initiative.

Introduction

Adoption of electronic medical records is on the rise, due to the Health Information Technology for Economic and Clinical Health Act and meaningful use incentives. Simultaneously, numerous HIE initiatives provide data sharing flexibility to streamline clinical care. Due to the consolidated data availability in centralized HIE models, conducting syndromic surveillance using locally developed systems, such as GUARDIAN, is becoming feasible. During the past year, Chicago has embarked on a city-wide HIE deployment campaign. Perhaps the most unique aspect of this endeavor is that the data warehouse for the HIE is intricately tied to the GUARDIAN syndromic surveillance system.

Methods

The GUARDIAN surveillance system has been running continuously at Rush University Medical Center (RUMC) for the past six years. In order to support real-time processing and analysis, the components of the system were deployed over six servers within the RUMC data center, specifically four processing servers plus two database servers configured as a single failover cluster.

Higher level (HL7) messages were received through transmission control protocol (TCP) connections and stored in a database-backed work queue. Using multiple servers, these messages were processed through a series of stages, specifically HL7 parsing, patient de-identification and matching, Natural Language Processing (NLP) of free-text, comparison of new patient data to stored disease profiles, report generation and user interaction through a web-based user interface. Based on the load metrics over the past six years, we have been able to scale up to a twelve server deployment (ten processing servers plus two database servers configured as a single failover cluster) which will support up to 30 hospitals within metropolitan Chicago.

Based on our experience, below is the list of challenges, solutions, and lessons learned:

1. Data structure to support rapid analysisa. Data stored in hierarchical tree-structure
   a. Free-text components of the patient chart are indexed by their unique concepts (as defined by the National Library of Medicine's Metathesaurus)
   b. Sophisticated NLP system to analyze and index free-text
2. Identification/ De-identification
   a. For patient data, cryptographically secure one-way hash (salted SHA256) was used
   b. Randomly-generated internal patient identifier to link HL7 messages
   c. For public health, re-identification can be carried out using the hash of the patient identifier and as much identifying information as is stored (e.g. the age and gender of the patient, the rough admit time and date, and the chief complaint)
3. On-boarding hospitals
   a. HIE integration less than 4 hours for majority of hospitals
   b. Redirect existing HL7 messages
   c. Compatibility issues with the site-to-site Virtual Private Network (VPN) and modes of TCP application. Thus, open additional ports than are typically needed over the site-to-site VPN and set up the TCP receiver services to operate in all modes
4. System performance metrics
   a. Though the average volume of messages increased from 26,000 to 45,000 per day, the average processing time per HL7 message from parsing to application of NLP remained similar (24 vs. 18 minutes)

Conclusions

GUARDIAN was able to demonstrate scaling up from two to seven hospitals with similar performance measures in terms of on-boarding and message processing times. The GUARDIAN system, and its associated data warehouse, was successfully expanded from one which supported a single group of hospitals, to one which can potentially support the hospitals of a major city.

Keywords

syndromic surveillance; GUARDIAN; Health Information Exchange