Surveillance in the Cloud: A New Platform for Disease Search and Situational Awareness

David W. Bauer[1], Ph.D., Mojdeh Mohtashemi[2, 3], Ph.D.

[1]High Performance Technologies, Inc.; [2]MITRE; [3]MIT CS and AI Lab

OBJECTIVE

We present a Cloud Computing based approach to disease surveillance that facilitates efficient data collection, processing and storage, as well as new concepts for data sharing and data fusion, disease search and situational awareness.

BACKGROUND

Major challenges in syndromic surveillance today include lack of standardization in syndrome definitions and limited ability to detect outbreaks of specific and rare diseases. To generate situational awareness surveillance results across various regions must be comparable and epidemiologically well defined. In addition, the high cost of obtaining and maintaining powerful computing resources (e.g., parallel computers) needed for data processing and analysis [1], and absence of a protocol for data sharing, highlight some of the obstacles to achieving situational awareness.

Cloud computing is an enabling technology that can overcome these challenges and facilitate new and novel approaches to surveillance.

METHODS

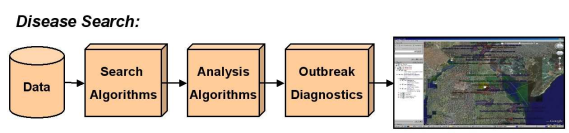

Figure 1. Conceptual diagram for the disease search platform.

Cloud computing is an emerging approach to distributed or grid computing that is generating a great deal of innovation in the computing industry. Cloud computing environments commonly provide two main components, computing and storage, both of which focus on providing high throughput computation for applications and I/O for data access. We refer the interested reader to [2] for additional information.

Our approach consists of several new methods for surveillance: disease search, cloud computing for analysis and a publish-subscribe model for data collaboration and fusion. In accordance with Figure 1, we define the term *disease search* to collectively refer to the computationally intensive process by which patient chief complaint data are grouped and optimized, detection algorithms are applied, and outbreak diagnostics are performed. Syndromic surveillance is a special instance of disease search, where the process for grouping data is limited to syndromic definitions rather than those of specific diseases.

The first component in our platform is a data fusion model for publishing and subscribing to data across health care entities (HCEs). Our platform constructs a cloud computing environment across existing computers within a HCEs IT infrastructure, and securely connects to other HCE clouds using *secure socket layer*. HCEs are then able to efficiently bind multiple forms of data to the cloud storage service and, using a publish-subscribe data model, share data with other HCEs and government agencies (e.g. CDC).

In the second component we perform an initial search algorithm based on a permutation of chief complaint data. We compute the power set over the set of all possible values contained in the data. For each subset, we query the data for the logical union of the subset elements, and compute outbreak detection algorithms in the cloud computing environment in real-time.

In the final component, if a significant signal is identified by the analysis component, we compute an online, population-based differential diagnosis. Our system detects the presence of outbreaks for specific diseases, rather than a suite of diseases encapsulated within some non-standard, syndromic definition (e.g., ILI), thereby generating unbiased situational awareness.

RESULTS

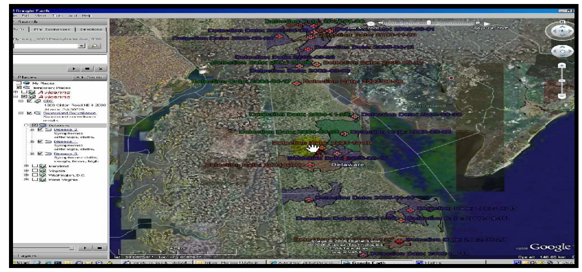

Figure 2. Multiple disease outbreaks in the Mid-Atlantic region.

Situational awareness is the natural extension of the set of interdependent operations of the aforementioned components, integrated within a single graphical interface such as Google Earth. Figure 2 represents an example capturing the results from executing our disease search platform. For patient confidentiality, the original data consisting of zip codes and time of visit were randomized. The set of chief complaint data used here are those of ILI syndrome provided by a public health department. For analysis, we applied Kulldorff's space-time permutation using SatScan software [3]. The data model in Figure 2 simulates a regional perspective scenario in which space-time data for multiple local- and state-level HCEs around the mid-Atlantic region over a 4-year period are used to generate situational awareness. Individual outbreaks are uniformly color-coded, and users may click on any spatial area (e.g., zip code) to display the SatScan analysis statistics as well as the outbreak-related differential diagnostics.

The platform executes detection algorithms on subscribed chief complaint data in real-time, and publishes the surveillance results for the mid-Atlantic region. Computing resources across HCEs cloud computing environments are effectively used in the computation, and the results are stored in the cloud storage service for consumption by other HCEs.

CONCLUSIONS

In this paper we utilize cloud computing as a low-cost computational platform to address critical challenges and limitations in today's surveillance systems and introduce novel concepts to facilitate disease search and situational awareness. We are currently evaluating our platform for disease search according to the guidelines proposed in [4]. Finally, this approach is of great value for early diagnostics of symptomatic data in emergency situations where the initial symptoms are often indicators of rare diseases (e.g., Sept 11th and Hurricane Katrina), as well as in resource-constrained regions of the world, where the high cost of computational resources and absence of skilled personnel are among many obstacles in disease tracking and control.

REFERENCES

[1] Bauer D, Higgs B and Mohtashemi M (2007) High Performance Computing for Disease Surveillance. Lecture Notes in Computer Science 4506:71.

[2] Cloud computing. (2008, July 17). In *Wikipedia, the free encyclopedia*. Retrieved July 17, 2008, from http://en.wikipedia.org/w/index.php?title=Cloud_computing&oldid=226274263.

[3] Kulldorff M, Heffernan R, Hartmann J, Assuncao R, Mostashari F (2005) A space-time permutation scan statistic for disease outbreak detection. PloS Med 2(3):e59.

[4] Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V (2004) Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks. MMWR 53(RR05):1-11.

Further Information: David Bauer, dbauer@hpti.com