IEEE VAST Challenge 2011

Detailed Task Description for all Challenges

Read this first

Mini-Challenge 1 - Characterization of an Epidemic Spread

Vastopolis is a major metropolitan area with a population of approximately two million residents. During the last few days, health professionals at local hospitals have noticed a dramatic increase in reported illnesses. Observed symptoms are largely flu-like and include fever, chills,sweats, aches and pains, fatigue, coughing, breathing difficulty, nausea and vomiting, diarrhea, and enlarged lymph nodes. More recently, there have been several deaths believed to be associated with the current outbreak. City officials fear a possible epidemic and are mobilizing emergency management resources to mitigate the impact. You have been charged with providing an assessment of the situation.

We provide you with two datasets. The first one contains microblog messages collected from various devices with GPS capabilities. These devices include laptop computers, handheld computers, and cellular phones. The second one contains map information for the entire metropolitan area. The map dataset contains a satellite image with labeled highways, hospitals, important landmarks, and water bodies. We also provide supplemental tables for population statistics and observed weather data. Additional information is provided in the README file.

MC 1.1 Origin and Epidemic Spread: Identify approximately where the outbreak started on the map (ground zero location). If possible, outline the affected area. Explain how you arrived at your conclusion. (Short answer)

MC 1.2 Epidemic Spread: Present a hypothesis on how the infection is being transmitted. For example, is the method of transmission person-to-person, airborne, waterborne, or something else? Identify the trends that support your hypothesis. Is the outbreak contained? Is it necessary for emergency management personnel to deploy treatment resources outside the affected area? Explain your reasoning. (Detailed answer)

Mini-Challenge 2 – Computer Networking Operations at All Freight Corporation

All Freight Corporation is a U.S. company that provides a range of shipping services, focusing on long haul and large-inventory commercial cargo. All Freight started as a regional shipping provider but over time has expanded to providing shipping services throughout the entire United States. Whereas expansion plans call for expanding their transportation options to include air transportation, All Freight currently ships freight exclusively via trucks.

To support its business, All Freight operates a corporate computer network. A computer network operations (CNO) group at All Freight is responsible for managing all aspects of the corporate network. This includes system administration tasks, so that All Freight can conduct business (e.g., managing customer accounts, operating a web interface that allows customers to book orders directly, and scheduling truck routes). Lately, with increased public awareness of cyber attacks, the CEO of All Freight has also asked you, as technical lead of the CNO, to improve the overall situation awareness of the corporate network. The CEO thinks that situation awareness will help in managing daily operations as well ensuring computer security. This new cyber situation awareness project will be an added duty for the CNO team; no new hires are authorized to cover additional workload. As the technical lead, you are tasked to develop a situation awareness interface that will give insight as quickly and clearly as possible in order to minimize the burden on the CNO team. The interface should integrate essential information to enable comprehension at a glance.

These datasets are provided as possible input to the situation awareness interface:

* A file describing the computer network architecture – This documentation includes a list of priority computers. These high priority nodes are essential to All Freight's ability to conduct business.
* Security policy rules
* A firewall log
* An IDS log

* An aggregated file of syslogs for all the hosts on the network
* A Nessus Network Vulnerability Scan Report

MC 2.1 Events of Interest: Using the new situation awareness display(s), what noteworthy events took place for the time period covered in the firewall, IDS and syslog logs? Which events are of concern from a security standpoint? Limit your answer to no more than five noteworthy events. For each event, at least one of the submitted screen shots must be relevant in your explanation of the event. (Detailed Answer)

MC 2.2 Timeliness: For each event submitted in MC 2.1, how early in the course of the event would your display(s) enable a CNO team member to recognize that the event was noteworthy? For each event, specify the earliest moment of recognition as a timestamp and provide a screen shot at the earliest moment of recognition. Explain how the CNO team member had enough information to determine that the event warranted attention. (Detailed Answer)

MC 2.3 Recommendations: What are the implications of the events discovered in MC 2.1? What report should the CNO give to the CEO and/or what actions should the CNO take to improve security? (Short Answer)

Mini-Challenge 3 - Investigation into Terrorist Activity

Intelligence analysts are looking for information related to potential terrorist activity in the region.

We provide you with a text corpus containing news reports. Each news report is a plaintext file containing a headline, the date of publication, and the content of the article.

MC 3.1 Potential Threats: Identify any imminent terrorist threats in the Vastopolis metropolitan area. Provide detailed information on the threat or threats (e.g. who, what, where, when, and how) so that officials can conduct counterintelligence activities. Also, provide a list of the evidential documents supporting your answer. (Detailed answer)

Grand Challenge - Cause and Effect

In Mini-Challenge 1, you used microblog data to characterize an epidemic spread. In Mini-Challenge 2, you conducted cyber security analysis for situational awareness of a corporate network infrastructure. In Mini-Challenge 3, you investigated terrorist activity in the region.

For the Grand Challenge, you are charged with investigating the cause of the epidemic.

In particular, you need to address the following:

Are any terrorist activities related to the current epidemic?
Describe the series of events, planned or otherwise, that led to the current epidemic.

Recall that we are particularly interested in how the interactive visualizations helped you with your analysis. Remember to explain in your process what visualizations you used and what insights were gained from these visualizations.

No additional data set is needed for the Grand Challenge - only the data provided in mini-challenges 1, 2 and 3.

For the grand challenge you need to provide:

- A debrief
- A Video showing how you conducted the analysis
- An optional Two-Page Summary

Questions? See the discussion blog or send email to challengecommittee AT weblab.cs.uml.edu

The 2011 VAST Challenge consists of three Mini Challenges and a Grand Challenge. Contestants can choose to work on one, some, or all of the challenges. To successfully respond to the Grand Challenge, contestants must tie

together all data sets with an overall scenario description using data elements from each of the three mini challenges, but are not required to submit to the mini-challenges.

The datasets used for these challenges are synthetic: that is, they are a blend of computer- and human-generated data. All datasets, whether real or synthetic, have anomalies. Some anomalies may be significant, some may not. Any anomalies reported should be supported by the proposed hypotheses. For example, “all first names start with a ‘M” may be interesting, but unless it is tied to the discussion of the situation, that anomaly has no place in your submission.

We will include all information necessary to form working hypotheses for the purpose of these challenges. No external data is needed to successfully perform the analysis. Be aware that using additional non-provided data may skew an otherwise successful solution.

Definitions

There are different formats/size for providing answers to the questions:

Short Answer:

Short answers are only requested in the mini challenges.
A short answer is a text description of the answer and of how you arrived at the answer. It is limited to 150 words (including captions) and a maximum of 2 screen shots.

Detailed Answer:

Detailed Answers may be requested by the mini challenges and the Grand Challenge.
A Detailed Answer is a longer text description focusing on how you arrived at the answer with much more details than the Short Answer.
• For mini challenges, detailed answers are limited to 1000 words (including captions), with a maximum of 5 screen shots.
• For the Grand Challenge there is no size limit (but less than 5000 words is recommended with a maximum of 15 screen).

Detailed answers should provide the answer and describe in detail the PROCESS USED TO ARRIVE AT THE ANSWER.

Please check the Guidelines page before you prepare your answers.

Video:

All entries are required to include a video with voice narration.
Maximum length (shorter is better):
• 4 minutes for Mini Challenge entries
• 15 minutes for Grand Challenge entries.
NOTE: If you submitted an entry to all three mini challenges you already have three videos for them. You may reuse all or parts of the 3 videos but should also leave enough time to show how you integrated the multiple datasets and come up with the grand challenge answers.

Please check the Guidelines page before you prepare your video.

Debrief:

Debrief are requested only in the Grand Challenge. The debrief is basically the analytic product that a professional analyst would deliver after doing the analysis.

A debrief is a maximum of 2000 words narrative describing your hypothesis about the situation at hand. Include in your narrative the relationships of the various players. If there are uncertainties, you can suggest possible next steps to clarify those uncertainties.

Please check the Guidelines page before you write your debrief.

Two-Page Summaries:

Two page Summaries are OPTIONAL, and they do not need to be submitted until after the results have been announced. They appear in the printed materials of the Symposium and also archived online.

These summaries allow you to give a general overview of your tools, significantly highlight novel features, provide references to papers and other relevant work and describe any new discoveries you made about your tools while working through the Challenge problem. Only the two-page summaries of the best entries (which are awarded an award) will be published in the Proceedings. Nevertheless, ALL submitted two-page summaries will be published online - along with your answer - in the VAST Benchmark Repository, whether or not they received an award.

The two-page summary should be formatted according to the general IEEE VGTC Guidelines

http://www.cs.sfu.ca/~vis/Tasks/camera.html
