IEEE VAST CHALLENGE 2009

Sample Data

- HOME
- DOWNLOAD
- TASK DESCRIPTION
- CRITERIA FOR JUDGING

- HOW TO SUBMIT?
- ANSWER FORMS
- RESULTS
- STUDENT SUPPORT

- DISCUSSION BLOG
- HISTORY OF CHANGES

VAST 2009 SAMPLE DATA (POSTED MARCH 4TH 2009)

The VAST 2009 Challenge is centered on a cyber analytics scenario: an employee is leaking important information to the outside world; hypotheses about his identity and network need to be made or confirmed.

It will include three mini challenges. Each mini challenge can be analyzed individually, while the Grand Challenge will require you to analyze and combine the data from all 3 mini challenges. Additional background information concerning the situation and circumstances of interest will be also included (a few pages of text).

Below are descriptions and example of the data we will provide later on in March.

Mini Challenge 1 (Badge and Computer Network Traffic)

Two datasets will be provided:

1) Facility access information: The data is about employees entering and leaving a facility, where they must use their badge (also called “proximity card” or “prox card”) to gain access to either the building or a classified area inside the building. The prox card data looks like this:

User Warning	Datetime	ID	Type
Synthetic Data	2008-01-01T06:35	4	prox-in-building
Synthetic Data	2008-01-01T06:54	36	prox-in-building
Synthetic Data	2008-01-01T07:11	32	prox-in-building
Synthetic Data	2008-01-01T07:21	40	prox-in-building
Synthetic Data	2008-01-01T07:45	51	prox-in-building
Synthetic Data	2008-01-01T07:49	9	prox-in-building
Synthetic Data	2008-01-01T07:49	20	prox-in-building
Synthetic Data	2008-01-01T07:54	43	prox-in-building
Synthetic Data	2008-01-01T08:01	55	prox-in-building
Synthetic Data	2008-01-01T08:03	51	prox-in-building
Synthetic Data	2008-01-01T08:05	29	prox-in-building
Synthetic Data	2008-01-01T08:06	51	prox-in-building

The data will be provided as a tab-delimited table with values of a generic warning, the event datetime, the employee id, and the type of event (prox-in-building, prox-in-classified, prox-out-classified). The field with “Synthetic Data” as a value is just a reminder that this is artificially created information.

2) The second dataset has computer use data of the people who work at the facility, in the form of IP logs. The IP log data looks like this:

User Warning	ReqSize	RespSize	Socket	SourceIP	AccessTime	DestIP
Synthetic Data	453	4026	8080	55.170.100.26	2008-10-01-08:00:01:188	83.112.96.249
Synthetic Data	824	5575	80	55.170.100.34	2008-10-01-08:00:11:594	195.41.197.95
Synthetic Data	2388	4448	80	55.170.100.3	2008-10-01-08:00:45:599	125.167.203.69
Synthetic Data	552	5714	8080	55.170.100.17	2008-10-01-08:00:58:943	34.106.41.110
Synthetic Data	1371	5316	80	55.170.100.38	2008-10-01-08:01:39:871	177.96.56.190
Synthetic Data	314	7100	80	55.170.100.13	2008-10-01-08:01:55:370	245.17.68.57
Synthetic Data	1632	5232	80	55.170.100.8	2008-10-01-08:01:55:754	55.170.30.100

Synthetic Data	640	8106	8080	55.170.100.27	2008-10-01-08:02:09:910	177.96.72.86

This will be provided as a tab-delimited table containing the sizes of the request and response, the source IP address, and the destination IP and port.

Mini challenge 2 (Social Network with Geospatial)

The data describes a social network, with three pieces of information: person to person links, person to geographic location links, and a map.

There will be two tab-delimited tables, one describing entities, i.e., either user names, city or country, and one containing links. As we have only user-to-user connection information, please consider these connections two-way links.

Finally, we have a map showing a country, its major cities, and information about neighboring countries and their major cities.

Format and examples:
A table of entities (listing all user names and places. We chose to add a "@" before the user names to avoid confusion with actual names):

ID	**User Name**	**Type**
INTEGER	STRING	STRING
1	@Arthur	Person
2	@Jerry	Person
3	Kannvic	City
4	Flovania	Country

And a big table of links:

ID1	**ID2**
INTEGER	INTEGER
1	2
2	22
2	3
100	1002
55	57
55	4

The map will be provided as a .jpg file.

Mini Challenge 3 (Video)

The data consists of video footage taken by a security camera for event analysis. It will be provided in Quicktime format. Sample video.

vast_clip_SAMPLE.zip

Grand Challenge
Some supplementary information will be provided (a Word Document, about one or two page long).