

Supporting Chat Exploitation in DoD Enterprises

Christopher D. Berube (Presenter)

Janet M. Hitzeman, Ph.D

Roderick J. Holland

Robert L. Anapol

Stephen R. Moore

The MITRE Corporation

June 19-21, 2007

Problem Statement

- **Use of internet chat in DoD enterprises is pervasive**
 - Widespread use in Air Operations Centers (AOCs)
 - Making its way onto airborne platforms (e.g., AWACS, JSTARS)
- **Operators gain awareness by monitoring chat rooms**
 - Activity in the chat rooms gives viewpoints on the current situation
 - Often many chat rooms at a time must be monitored
- **An operator must not miss anything important**
 - No time for breaks
 - No allowance for inattention
- **Chat exploitation needs vary across DoD enterprises**
 - **Alerts** on “interesting” events in chat may be required for tactical operations
 - **Summaries** of events in chat may be required at command levels

Need tools to help a broad set of users better manage and exploit military chat

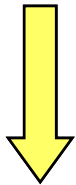
MITRE Technology Program Research Project

- MITRE's Technology Program funds internal research projects through its sponsors, e.g., the Air Force
 - Projects with a near-term, customer-specific focus are known as **Mission Oriented Investigation and Experimentation (MOIE)** projects
- Today, we present the results of a second-year Air Force MOIE project, “Facilitating Sense Making for Situational Awareness”
 - Focus is on the management and exploitation of military chat for a broad set of users within DoD
- Project objectives
 - Apply MITRE's Alembic information extraction system to the problem of **extracting tactically relevant data from military chat**
 - Develop a set of tools to help the user best exploit Alembic's products
 - User interfaces
 - Functionality for retrospective analysis and summarization of chat
 - Develop a software prototype to demonstrate our research
 - Chat Exploitation Tool (ChET) is our laptop-based demo

Technical Approach – Extracting Entity Class Data from Chat

[12:24] <Exec_Dir> CCO-where's Jake? He's heading to Medford - and we swing him straight there?
[12:26] <Exec_Dir > We're thinking Jake's going to be faster than pulling Trapper 11 from Gat D. However, we are checking with TF-11 if we can pull Trapper.
[12:26] <HQ_CCO> Yes. We can move Jake that way now.
[12:26] <HQ_CCO> Rough estimate 20-30 min
[12:27] <Exec_Dir > Great – Jake's on the way.
[12:28] <Boss_OPS> We've got **Trapper 11** in Gat d and **GCAS** ready for **launch**
[12:30] <Exec_Dir > CCO Boss –we got an airpower gap when Jake departs at 1300 for 1 hr. May need GCAS to fill the GAP if this goes that long.
[12:30] <Exec_Dir > CCO - 15 NM from the border – please call Bulldog.

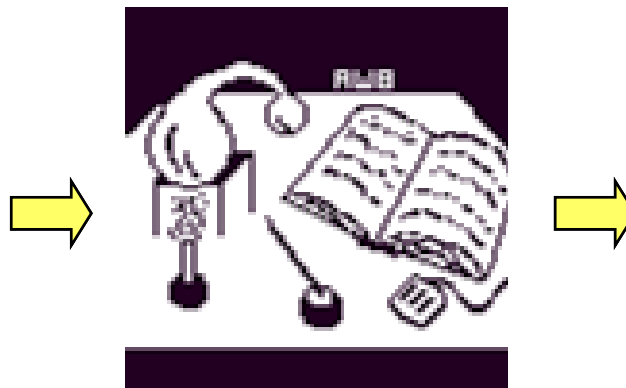
10 minutes of chat from a single chat room



A military language pre-processor (MLPP) uses a regular expression to *extract an aircraft voice call sign* (e.g., Misty 11):

"([a-zA-Z]{4,12})(\s?\d{1,2})"

Alembic Information Extraction



Extracted <AirMission> Class

<AirMission>		
	<ACType>	
	<ACCallSign>	
	<MissionType>	
	<MissionNumber>	
	<ATONumber>	
	<State>	
		<Location>
		<ZTime>
	<Activity>	
		<Location>
		<ZTime>

How Alembic Extracts Entity Class Attributes

■ Wordlists

- <AirMission><ACType>: B-52
- <Missile><Name>: Patriot

■ Regular expressions (i.e., **rules**) with help from context

- “...**msn#**8462AC...”

■ Context alone

- “...**bty** 1124 is ready to...”

Shorthand for “battery” helps determine that what follows is a air defense unit

- This is an <AirDefense><Unit>, not a <AirMission><MissionNumber>

- “...**Scud** launch is expected IVO...”

- “launch” refers to a <Missile><Activity>, not an <AirDefense><Activity>
- Here, words (e.g., “Scud”) tagged first by the MLPP provide context

■ Also considers

- Different form/same meaning: Fire/fired/firing/fires, unit 436/114
- Telegraphic speech: “...missile ivo camp...”

Alembic information extraction models are being developed for **ELINT, AirTarget, AirMission, Missile, Location, MASINT, ErrorEllipse, AirDefense and GroundTarget** entity classes

Entity Class “White Papers”

■ Provides an unclassified set of lists, rules, contexts for the development of Alembic information extraction models

- <Air Defense>, <AirMission>, <ELINT> and <Missile> entity classes

<AirMission><MissionNumber>:

ATO format [1-8ANBS] **AND** preceded or followed by

msn
msn#
mission#
mission num
mission number
msn#s

<AirMission><State>:

on station, on-station
on (the) deck, on-deck
off station, off-station
airborne

<AirMission><Activity>:

launch, launched, launches, launching
escort, escorted, escorts, escorting
refuel, refueled, refuels, refueling
recover, recovered, recovers, recovering

<AirMission><ACType>:

Designator	Name	Nickname
B-52H	Stratofortress	
B-1B	Lancer	
B-2	Spirit	
S-3B	Viking	
P-3C	Orion	
P-7		
AH-1W	Super Cobra	
A-4	Skyhawk	
OH-6A	Cayuse	
AH-6J	Little Bird	
A-6E	Intruder	
EA-6B	Prowler	
A-7	Corsair II	
AV-8B	Harrier	
A-10	Thunderbolt II	Warthog, Hog
OA-10	Thunderbolt II	Warthog, Hog
A-12	Avenger II	
AH-64A	Apache	
AH-64D	Apache	Longbow , LBA
RAH-66	Comanche	
F-111		
FB-111		

<AirMission><MissionType>:

Mission Data Code	Mission Type Name
CAS	CLOSE AIR SUPPORT
GCAS	GROUND ALERT CLOSE AIR SUPPORT
XCAS	AIRBORNE ALERT CLOSE AIR SUPPORT
AI	AIR INTERDICTION
ATK	ATTACK
AH	ATTACK HELICOPTER
BAI	BATTLEFIELD AIR INTERDICTION
SEAD	SUPPRESSION ENEMY AIR DEFENSE
ATK	ATTACK
ILLUM	FLARE ILLUMINATION

Alembic Performance

- The information extraction performance of Alembic is measured against a set of test chat logs, separate from the training set, using the F-score measure
- The **F-score** is a measure of how well the system did, weighted by Precision and Recall
 - **Precision** takes false positives into account, i.e., the degree to which the model applies incorrect tags
 - **Recall** takes false negatives into account, i.e., the degree to which the model doesn't tag a phrase that should be tagged
- Results for two chat rooms
 - Combined Forces Air Component Commander (CFACC)
 - Combined Forces Land Component Commander (CFLCC)

Chat Room									
	Entity Class	# Correct	# of Tags	% Correct	# False Positive	# False Negative	Precision	Recall	F-score
CFACC									
	<AirDefense>	336	437	76.9	38	63	0.90	0.84	0.87
	<AirMission>	17	30	56.7	6	7	0.74	0.71	0.72
	<ELINT>	129	157	82.2	6	22	0.96	0.93	0.94
	<Missile>	31	48	64.6	4	13	0.89	0.70	0.78
CFLCC									
	<AirMission>	163	212	76.9	25	40	0.85	0.79	0.82

Chat Exploitation Tool (ChET) – UserDisplay GUI

The screenshot displays the UserDisplay GUI with the following sections:

- Entity Classes:** A list of entity types including <AIRDEFENSE>, <AIRMISSION>, <AIRTARGET>, <ELINT>, <ERREORELLIPSE>, <GROUNDTARGET>, <LOCATION>, <MASINT>, and <MISSILE>.
- Chat Log:** A list of chat messages with timestamps and content. The selected message is: [Kmart] [SIG_MAN] [2006-03-24 03:25:00] Follow-up on <ELINT> <ELNOT> **A123Z** </ELNOT> </ELINT> contact: TA/TT <ELINT> <TYPE> **radar** </TYPE> </ELINT> , thought to be <ELINT> <STATE> **non-operational** </STATE> </ELINT> at <LOCATION> <LATITUDE> **422812N** </LATITUDE> </LOCATION> <LOCATION> <LONGITUDE> **0711647W** </LONGITUDE> </LOCATION>.
- Detail for selected chat alert:** A detailed view of the selected chat message, showing the full text and the entity classes used in the message.
- All chat in rooms specified by user "Chris":** A list of chat messages from the selected room, showing the full text and the entity classes used in the message.
- Correlated chat utterances across chat rooms:** A section showing the results of a 2D Spatial Correlation analysis, with a threshold of 10 nmi. It lists chat message groups and the messages within them.

List of chat alerts

Detail for selected chat alert

All chat in rooms specified by user "Chris"

Correlated chat utterances across chat rooms

ChET – UserProfile GUI

List of chat rooms

List of entity classes

Filter (list, time) for aircraft call sign

The screenshot displays the 'UserProfile' window with the following sections:

- Chat Rooms:** A list of chat rooms with checkboxes for CFACC, Kmart, and ISR, all of which are checked.
- Entity/Event Class Data in Chat:** A grid of checkboxes for entity classes: ELINT, MASINT, ErrEllipse, AirDef, AirMsn, Location, Missile, AirTarg, and GndTarg. ELINT, Location, and Missile are checked.
- AirMsn: AC CallSign:** A section with radio buttons for 'All', 'None', and 'Specific'. The 'All' option is selected. Below are input fields for 'Start DateTime(Z):', 'End DateTime (Z):', and 'AirMsn: AC CallSign:' with an 'Enter' button.
- Latitude/Longitude:** A section with radio buttons for 'None' and 'Specific'. The 'Specific' option is selected. Below is a large area titled 'Spatial, time filters' containing input fields for 'Start DateTime(Z):', 'End DateTime (Z):', 'Ctr Lat (DDMMSS):', and 'Ctr Lon (DDMMSS):'.
- Cross Chatroom Correlation:** A section with a 'Delta Time(sec):' input field and radio buttons for 'Rectangular Acceptance Region', 'Circular Acceptance Region', and 'Statistical Distance'. The 'Rectangular Acceptance Region' option is selected, with input fields for 'Delta Lat (deg):', 'Delta Lon (deg):', and 'Radius (NM):'. A 'Probability(%)' input field and an 'Enter' button are also present.
- Location: BE Number:** A section with radio buttons for 'All', 'None', and 'Specific'. The 'None' option is selected. Below are input fields for 'Start DateTime(Z):', 'End DateTime (Z):', and 'Location: BE Number:' with an 'Enter' button.
- ELINT: ELNOT:** A section with radio buttons for 'All', 'None', and 'Specific'. The 'All' option is selected. Below are input fields for 'Start DateTime(Z):', 'End DateTime (Z):', and 'ELINT: ELNOT:' with an 'Enter' button.
- Chat Keyword:** A section titled 'Keyword search' with input fields for 'Start DateTime(Z):', 'End DateTime (Z):', and 'Chat Keyword:' with an 'Enter' button.
- Profile Name:** A text field containing 'Chris' and a 'Commit' button.

The taskbar at the bottom shows the user is logged in as 'root' on a machine named 'mm136606-pc-1/rot', with a task titled 'MenuDemo'.

ChET – QueryPanel GUI

■ GUI for “drill down” analysis of chat content and participants

- Drop-down menus will provide list of available column values

- e.g., List of entity class names currently in chat database

■ Provides a retrospective/forensic capability

The screenshot shows the ChET QueryPanel GUI. The top window, titled "Query Display", allows users to configure a query. It includes a table for selecting columns and a "Run" button.

Include in Query	Class	Attribute	Chat rooms	User ID	From	To
<input checked="" type="checkbox"/>	<AIRMISSION>	<ACCALLSIGN>	CFACC	any	2006-03-24 01:00	2006-03-24 21:00
<input checked="" type="checkbox"/>	<ELINT>	any	any	any		
<input type="checkbox"/>	any	<ACTYPE>	any	any		
<input type="checkbox"/>	any	<ACCALLSIGN>	any	any		
<input type="checkbox"/>	any	<MISSIONTYPE>	any	any		
<input type="checkbox"/>	any	<MISSIONNUMBER>				
		<ATONNUMBER>				
		<ACTIVITY>				
		<STATE>				

Below the configuration window is a "Run" button. The bottom window displays the query results in a table:

DateTime	ChatRoom	UserID	Chat Message
2006-03-24 03:26	CFACC	Exec_Dir	We're thinking Jake's going to be faster than pulling Trapper 11 from Gat D. However, we are checking with TF-11 if we can pull Trapper.
2006-03-24 03:27	CFACC	ELINTER	Al Azah spoonrest is active at 0325Z
2006-03-24 03:28	CFACC	MISS_OPS	We've got Trapper 11 in Gat d and GCAS ready for launch
2006-03-24 03:28	CFACC	MISS_OPS	BTW, Trapper 11 is on msn# 1234
2006-03-24 03:31	CFACC	ELINTER	Al Azah spoonrest at 422812N 0711647W
2006-03-24 03:32	CFACC	Exec_Dir	Trapper 11 staying in Gat D as fraggd.
2006-03-24 03:33	CFACC	MISS_OPS	I've got Trapper 11 RTB from Gat D, would you like them?
2006-03-24 03:07	Kmart	SIG_MAN	ELINT contact: A123Z active IVO Trenton @ 0305Z
2006-03-24 03:25	Kmart	SIG_MAN	Follow-up on A123Z contact: TA/TT radar, thought to be non-operational at 422812N 0711647W
2006-03-24 03:27	Kmart	SIG_MAN	ELINT contact: EW radar up IVO Fort Lee

The bottom status bar shows the user is logged in as "root@mm136606-pc-1/us" and the application is "MenuDemo".

On-going Research – Extracting Event Classes

- Use Alembic’s facility for *inferencing* to combine “eventlets” in a single chat room that describe parts of the same full event
 - Heuristic: “Eventlets” are contained in a single chat message

[User_ID1] “Al Azah spoonrest is active at 1815Z.”

•
•
•

[User_ID2] “Al Azah spoonrest at 422812N 0711647W 1.3 x 0.2 164.4”

Here, two ELINT class attributes match, so combine “eventlets” into a single “event” with a union of class attributes.

<ELINT>	
	<Name>
	<Type>
	<ELNOT>
	<Location>
	<State>
	<ZTime>

<Location>		
	<Name>	
	<Latitude>	
	<Longitude>	

<ELINT>	
	<Name>
	<Type>
	<ELNOT>
	<Location>
	<State>
	<ZTime>

<Location>			
	<Name>		
	<Latitude>		
	<Longitude>		
		<ErrorEllipse>	
<ErrorEllipse>			
	<SMajAxis>		
	<SMinAxis>		
	<Orientation>		

On-going Research – Correlating Entities/Events Across Chat Rooms

- Correlating data across chat rooms
 - Compute similarity measure between pairs of entity/event class instances
- Numeric and non-numeric attributes

<AirDefense> entity class extracted from Chat room “A”

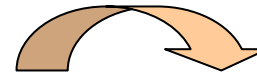
(Chat “object” #1)

<AirDefense>	
	<Name>
	<Type>
	<Location>
	<Activity>
	<State>
	<ZTime>
	<Unit>

<ELINT>	
	<Name>
	<Type>
	<ELNOT>
	<Location>
	<State>
	<ZTime>

<ELINT> entity class extracted from Chat room “B”

(Chat “object” #5)



Similarity Matrix

	1	2	3	4	5	6	7	8
1	1	0	0	0	0.9	0.3	0.55	0
2	0	1	0	0	0	0.85	0	0.4
3	0	0	1	0	0	0	0	0
4	0	0	0	1	0	0	0	0
5	0.9	0	0	0	1	0	0	0
6	0.3	0.85	0	0	0	1	0	0
7	0.55	0	0	0	0	0	1	0
8	0	0.4	0	0	0	0	0	1

Only show users pairs of Chat objects with similarity $\geq 90\%$

- Threshold (“hard” correlation) or cluster (“soft” correlation) groups of similar chat objects
- Initially, only location and time attributes are used to correlate pairs of chat objects
 - e.g., Distance between pairs of chat utterances containing latitude/longitude < 10 mi

Summary

- **MITRE's Alembic information extraction system has been applied to military-style chat**
 - Entity class data can be accurately extracted and tagged
 - Allows a means to index into chat
- **A research prototype has been designed around Alembic to serve a broad range of chat exploitation needs:**
- **Alerts/Monitoring –**
 - Real-time alerts on data types of interest, tailored by a set of user-specified preferences
- **High-level Reporting/Summaries –**
 - Querying the prototype's database containing indexed chat utterances by entity class and class attributes
 - Means to generate summaries of chat sessions by topic (entity class), chat room or user
- **Forensics –**
 - The ability to correlate extracted chat utterances across multiple chat rooms