Military Textual Analysis and Chat Research

Emily W. Medina

Space and Naval Warfare (SPAWAR) Systems Center Pacific
emily.medina@navy.mil

Abstract

Military use of chat, as a grassroots technology, has become so widespread in the last fifteen years that Joint communities have just begun to scratch the surface of this robust communications technology. Historically, IRC room-based chat has been the preferred paradigm for tactical users, due to its ease of use and robustness on high-jitter networks. Migration to XMPP-based chat is now being discussed for a way forward, though this change in protocol raises many questions for the current concept of operations (CONOPS), especially for edge, tactical, users.

This special session will focus on current research being performed within DoD to analyze current usage of chat, in order to assist Fleet users both today and in the future.

1. Introduction

Chat is part of a domain called computer-mediated communication (CMC). As a discourse medium, CMC exhibits characteristics that are distinct from standard spoken or written discourse: it is faster than other written communication, but slower than spoken exchange; it allows the participants to edit an exchange; and it allows for simultaneous exchange among multiple participants [6]. Computer-mediated discourse (CMD), a specialization within CMC, exhibits non-standard linguistic features apart from both written and spoken discourse, features that rely on the theory of "economy of effort" [6]. Acronyms, substitutions, mimicry of visual and spoken language features—called emoticons—and deliberate representations of prosody (non-linguistic features such as tone, i.e. YELLING) all have distinctive roles in CMD. CMD can be either asynchronous (email), or synchronous (chat, instant messaging), depending on whether users are present and active at the time communication is sent [6]. Internet Relay Protocol chat, or IRC, is somewhat unique in that a room can exist beyond the time users are logged in; users can join and exit a room at any time, and the conversation can continue with other participants. Chat rooms have dominated the Internet as a social medium since the early 1990s; over time, however, online communication has gradually moved to peer-to-peer/instant messaging services (IM). Chat room usage has instead reinvented itself in the use of online forums. Forums are far more structured in nature; often, they are hierarchically organized by subject, and are much more asynchronous.

The difference between IRC and IM goes beyond look and feel, however; these are distinct protocols that behave very differently from one another. In IRC, users can subscribe to chat rooms (channels) already in existence. Once a user subscribes to a channel and joins a particular room, his username is displayed for other members of the room to see. Any point-to-point communication is accomplished using *whispers*, an IM-like capability. However, whispers are not automatically logged, while most public chat text is stored on the server.

IM, on the other hand, is an XML-based protocol that primarily involves point-to-point communication primarily. Once a user logs into his XMPP client, available users are displayed as in chat; the major difference, however, is that displayed users may or may not "see" one another, depending on who has subscribed to whom. Some XMPP clients offer conference services that are chat room-like in nature; however, this is an add-on capability.

This paper will touch on various efforts within the Navy at approaching and augmenting tactical chat as it is used in theatre today. DoD-wide stances on chat as a collaboration initiative, Naval Postgraduate School (NPS) work on unclassified chat, and SSC Pacific work on tactical chat will be discussed in turn. Additionally, the initial development of a tactical chat corpus will be presented, with a look at future direction and research.

2. Military chat characteristics

Military chat usage essentially began during the Persian Gulf War in Iraq as a system administration troubleshooting tool [4]. Many warfighters who relied on radio communications found chat to be superior for consistent weather and unit report updates, due to it low overhead, cross-platform capabilities. Operation Iraqi Freedom saw chat usage skyrocket to thousands of users within IRC [8]. Communities within each command have unique concepts of operations (CONOPS) to conduct tasking over IRC, but as of yet there exists no overarching standard across Joint communities for chat operations. There is a push to move from IRC-based operations to XMPP, as XML provides many additional chat capabilities, including parsing and automated tools [3]. The Defense Information Systems Agency (DISA) currently has two XMPP chat tools under review, though no decisions have been made at this time towards a transition plan for IRC users [3].

Tactical chat differs in a few important ways from public chat on the Internet. There has been a rise in documentation of conventions and standards for conducting daily tasking within chat, mostly focused on appropriate subject matter and data input for certain channels. User IDs in tactical chat are based on functional roles, instead of marking the identity of a particular individual; it is possible that many warfighters can chat in a room using the same user ID. The value in using functional roles is that each chat room participant is clearly identifiable, and appropriate questions can be directed in a timelier manner.

Orders are given over chat rely using a specified format and procedure. Even though chat is being used more frequently than in the past, it is still not considered a standard venue for transmitting orders, and is usually followed by a formal order process. The rule of thumb for chat is the transference of time-sensitive, critical collaboration of a type that would otherwise overwhelm a radio network.

In a highly stringent chain-of-command environment, chat rooms offer one of the only modes of informal guidance for new recruits. Though efforts have been made to formalize discussions in chat, it is still possible to pass information without fear of recourse from higher-ranking officers. All participants are responsible for their participation in chat, and for monitoring information passed during their watch. The ability to "whisper" within chat rooms often acts as an aid to a new watchstander when he needs help or clarification from a more seasoned comrade. In this kind of ad-hoc environment, however, there are issues that arise. Often, acronyms (a standard in military-speak, though difficult for a newcomer to parse) can overwhelm communications, and actually create more ambiguity between chat users. It is not uncommon for users to ask for clarification on certain acronyms when receiving formal information, since many acronyms are identical in different user communities.

Typing errors have also caused a number of problems, especially with current search tool capabilities in IRC. Simple word searches won't pick out information that users need to reference, and the potential for error in typing is high. Speed in communication is also affected in this medium.

Watchstanders responsible for chat often monitor events by tiling chat rooms across a computer screen. Watchstanders have been known to monitor up to 20 chat rooms at once, which quickly degrades visual effectiveness without proper alerting mechanisms. Most IRC users, unfortunately, have poor access to advanced alerting mechanisms beyond notification of new chat entries or simple color coding. Worse yet, chat is monitored in real-time; therefore, should a watchstander be distracted or pulled away from his watchstation, vital information may be missed altogether.

The primary issue with tactical chat is the information loss in transfer from one user to the next. Often, a watchstander's tour of duty lasts at most a few years; in that time period, he becomes the leading expert for tracking topics of interest and important information for that channel and that community. However, once this individual is rotated out to another position, most likely a completely disconnected position from his former post, his expertise is often lost on the new watchstander taking his place. There is no automated tracking of topics of interest or events that can be saved off and automatically searched. In light of these standing issues, work has begun to address and to secure chat's future as a lasting and effective collaboration tool.

3. Tactical chat analysis

Public, unclassified chat traffic contains intriguing features that tactical chat does not. For instance, each public user must subscribe to an online chat server, and in doing so provide some rudimentary information about himself. This information is publicly available, though private information, e.g. first and last name, remains protected. Since most public chat takes place in a forum-like situation, enhanced user features, such as age group or gender, are also provided. Chat research at the Naval Postgraduate School (NPS) [1] [5] has focused on creating an unclassified corpus of chat data, one of the first of its kind. Parsing and

application of existing tagsets has yielded a unique body of data with which to more closely examine chat discourse analysis.

Until a few years ago, no military chat corpus existed with which to perform any similar discourse analysis or further research. A collection of tactical chat data has grown slowly since 2003, aided mostly by researchers logging in and recording chat usage as observers. Chat room channels from servers worldwide were chosen based on those Fleet watchstanders regularly monitored; researchers were then able to log on and record chat traffic for the purposes of knowledge management automation. A sizable and growing database of over 10 million words of American military English is now available in a MySQL database (Ver.4.1.07) for more in-depth analysis. Fields stored within this database include the following: messageID, timestamp, chat server, chat channel, userID, and chat entry. These data are currently unannotated for part-of-speech or syntactic features. This corpus is currently classified at the GENSER SECRET level.

4. Future directions

While this tactical chat collection needs to undergo a fair amount of parsing and tagging, it is still possible to use these data for rudimentary unstructured textual analyses. Several projects have applied statistical natural language processing (NLP) techniques to these data in order to generate some initial ideas on event extraction and topic thread detection work. The Tactical Situation Assessment Technologies (TSAT) project has used this collection to extract unigrams and bigrams for common themes and concepts present in chat today [7]. Further work on topic classification and thread extraction is currently being performed [1], since it has shown to be useful for other chat studies [2]. Additional anti-submarine warfare (ASW)-specific chat content has been analyzed for similar themes and concepts, with a larger focus of providing semantic search for geographical and ship track updates. Social network analysis has been applied to chat room dialogue in the past [9]; tactical chat would definitely benefit from this type of analysis, as it may reveal working relationships that can be better aligned for more timely decision-making. Other studies have applied software agents to IRC chat, in order to summarize each channel's content and make recommendations to new users [10].

Work will continue in defining and enhancing this body of tactical chat data, as well as the requirements and transition architectures that will define future chat work within DoD. IRC is currently in end-of-life status, and migration to XMPP is imminent, though no official transition has been proposed. Current and future analysis of chat will help to guide these decisions for transition, as well as to better equip warfighters with automated decision aids.

5. Acknowledgements

6. References

[1] P.H. Adams, C. Martell. "Topic detection and segmentation in chat", *IEEE-ICSC 2008 Proceedings*, Santa Clara CA, 4-7 AUG 2008, in press.

[2] J. Bengel, S. Gauch, E. Mittur, R. Vijayaraghavan, "ChatTrack: Chat room topic detection using classification", *Second Symposium on Intelligence and Security Informatics (ISI 2004) Proceedings*, Tucson AZ, 10-11 JUN 2004, pp. 266-277.

[3] L. Duffy, "DoD Collaboration and Chat Systems: Current status and way ahead", *IEEE-ICSC 2008 Proceedings*, Santa Clara CA, 4-7 AUG 2008, in press.

[4] L. Duffy, "Internet Relay Chat (IRC): How a grassroots technology became an integral part of operational Command and Control", SSC SD Biennial Review, San Diego CA, 2006.

[5] E.N. Forsyth, C.H. Martell. "Lexical and discourse analysis on online chat dialogue", *IEEE-ICSC 2007 Proceedings,* Irvine, CA, 17-19 SEP 2007, pp. 19-26.

[6] S. Herring, "Computer-mediated discourse". In D. Tannen, D. Schiffrin, and H. Hamilton's (Ed.s) Handbook of Discourse Analysis, Blackwell, Oxford, 2001, pp. 612-634.

[8] E.W. Medina, L. Duffy. "Tactical Situation Assessment Technologies (TSAT)", *IEEE-ICSC 2008 Proceedings*, Santa Clara CA, 4-7 AUG 2008, in press.

[9] E.W. Medina, S. Fugate, L. Duffy, D. Magsombol, O. Amezcua, G. Rogers, and M. Ceruti. "Next-Generation Tactical Situation Assessment Technologies (TSAT): Chat", *ICWE 2007 Proceedings*, Como Italy, 16-20 JUL 2007, pp. 526-532.

[10] V.H. Tuulos, H. Tirri, "Combining topic models and social networks for chat data mining", *IEEE/WIC/ACM Web Intelligence Conference 2004*, Beijing China, SEP 2004, pp. 206-213.

[11] N.W. Van Dyke, H. Lieberman, P. Maes, "Butterfly: A conversation-finding agent for Internet Relay Chat", *International Conference on Intelligent User Interfaces (IUI 1999)*, Los Angeles CA, 5-8 JAN 1999, pp. 39-41.