Introduction to Traffic Analysis[1]

BY LAMBROS D. CALLIMAHOS

~~*Confidential*~~

A basic exposition of the principles and techniques of traffic analysis.

GENERAL

Traffic analysis is defined as that branch of cryptology which deals with the study of the external characteristics of signal communications and related materials for the purpose of obtaining information concerning the organization and operation of a communication system. By means of traffic analysis valuable information can be derived concerning the enemy and his intentions, even without actually reading the texts of the intercepted messages; the solution and translation of messages are the functions of cryptanalysis and not traffic analysis.

Traffic analysis can yield a detailed knowledge and thorough understanding of a communications network; traffic analysis techniques involve, among others, the reconstruction of the nets and the determination of the methods of their operation, the solution of callsign and routing or address systems, the solution of frequency rotation systems, the identification and analysis of components of the message externals, the interpretation of radio procedure, the study of the distribution of cryptosystems, and the analysis of authentication systems. The results obtained from traffic analysis materially contribute to the following:

Intercept operations. Traffic analysis provides information such as call signs, frequencies, locations, and schedules pertaining to target enemy stations, thus assisting intercept stations in the accomplishment of their missions; and, in coordination with cryptanalytic and intelligence interests, traffic analysis assists in establishing the priorities for the interception of individual circuits.

Cryptanalysis. Traffic analysis furnishes assistance to cryptanalysis in many ways, depending upon the particular communications situation; this assistance includes information as to the identity and location of radio stations, information of cryptanalytic interest gleaned from enemy operators' "chatter," identification of possible stereotype or proforma messages from external characteristics of the traffic, and identification of isologs.

[1] This article is an extract from the forthcoming NSA text, *Military Cryptanalytics, Part II*.—Editor

Intelligence. The organization of a radio network and the manner in which messages are passed over this network reflect troop disposition, command relationships, and impending movements and preparations for military activity; therefore an analysis of net structure, traffic contacts and patterns, traffic volumes, and similar communications features, is of considerable assistance in building up a complete intelligence picture.

Security. The techniques developed by traffic analysis in the attack on intercepted enemy communications may also be applied to our own monitored signal communications in order to uncover possible weaknesses and to maintain high standards of communication security by preventing these weaknesses from developing in our communications.

There are three kinds of basic data used in traffic analysis as follows:

Intercept data, comprising all information supplied by the intercept operator, and consisting of the frequencies on which transmissions are heard, the time the transmissions are heard, intercept operator comments such as signal strength and audibility, "fist" characteristics of the target radio operator, and any peculiarities in the transmission or handling of the traffic that strike the intercept operator as being significant or out of the ordinary.

The transmission, comprising everything transmitted by the target radio operator, and including the initial call-up, the exchange of call signs, the traffic passed, the servicing incidental to the traffic being passed, the radio operators' chatter, and the signing off. Traffic consists of the message externals (i. e., the preamble and postamble, if any) and the message text proper. The externals comprise various items that facilitate the handling of the message, among which are the radio station number and perhaps a message center number or other reference numbers, the group count, routing and address information, precedence indicators, the file date and time, etc.; all this information is of considerable potential value in traffic analysis. The message text, if it displays patent cryptographic characteristics, can also be of use.

Collateral information, comprising any information, other than that derived from a study of intercepted communications, which may be of value in traffic analysis; e. g., captured documents, intelligence reports, etc. In addition, traffic analysis is aided by *communication intelligence collateral* such as direction-finding bearings, Morse operator analysis, plain language messages, and decrypted traffic.

In traffic analysis the details of each feature of the communications operations or structure are studied, followed by analysis of the inter-

relationships among these features, culminating in the reconstruction of an entire net together with all the details of its operation. At the start of a traffic analysis problem, little may be known concerning the target communications; it would first be necessary to segregate initially intercepted traffic into several major types or nets by means of cryptographic features, common operating characteristics, or other means. At this point the intercept stations are given general search missions over the entire range of radio frequencies to intercept desired types of transmissions. As traffic accumulates, fragmentary nets are diagrammed and analysis is begun on the transmission characteristics and on the message externals, with particular emphasis on the preamble components and on routing methods; research is performed on call signs, frequencies, schedules, procedure signals, external message numbers, routing indicators, and cryptographic features, resulting in the ultimate reconstruction of the complete net with all its pertinent details.

RADIO COMMUNICATIONS

Efficient radio communications are dependent upon (1) the physical laws for the transmission, and (2) the requirements imposed by the necessities for the establishment and maintenance of communications. The first consideration involves the frequencies and power used, and the second consideration embraces the details necessary for the communications themselves, such as the call signs, routing, message numbering conventions, and receipting and servicing of the traffic. These latter items may be varied or changed by direction of the communications authority either for convenience in handling traffic, or for purposes of secrecy, or both.

From the standpoint of traffic analysis study there are three main aspects of radio operations, as follows:

The operating data. These consist of the basic operating and functioning data of the net; e. g., the structure or form of the net, the frequencies, the call signs, and the schedules.

The radio transmission. This includes the particular Morse code used, the procedure signals employed, the order of elements of the transmission, and radio operators' chatter.

The messages. These include the message texts proper, together with the message preambles and postambles. The cryptographic features of the message texts, such as discriminants and message indicators, the type of cryptographic text (whether in letters or digits), and the length of the code groups, are all of considerable assistance in traffic analysis; plain-language messages are also exploited.

OPERATING DATA

Radio stations are linked together and organized into nets for the purpose of intercommunication; this organization follows definite

patterns, reflecting the command structures since the lines of communication must coincide with echelons of command in order to meet military communication requirements. In a particular grouping of stations the one serving the senior echelon is the station usually in charge of the subordinate stations; this station is called the *net control station* (abbr. NCS), and the others are called outstations. The control station is responsible for the supervision of transmissions, procedures, and circuit discipline. A typical net structure is shown in Fig. 1, below. Station 1 is the superior headquarters, with Stations 2, 3, and 4

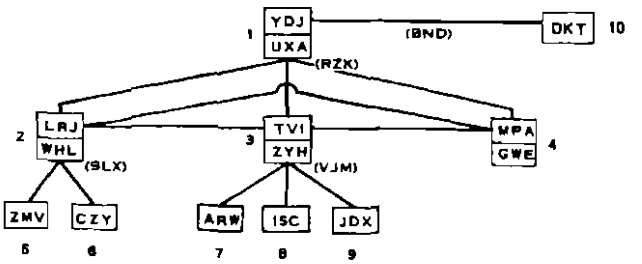

Fig. 1.

as the immediately subordinate outstations; Station 2 in turn has two outstations, and Station 3 has three outstations. Station 1 is also in communication with Station 10, the NCS of another net.

Stations are identified by one or more call signs which consist of a group of letters, digits, or both. In the diagram above, Stations 1–4 have two call signs each, while the remaining stations have but one call sign. Multiple call signs are used for convenience of operations, or for security; they are either in the form of *variant call signs* (the selection from these being left up to the radio operator) or of *split call signs* (the selection of the proper call sign being governed by the time of day, the radio frequency used, etc.).

The usual type of call-up is the double-station call procedure, wherein the call signs of the called station and of the transmitting station are sent, separated by the procedure signal DE (meaning "from"); for example, if TVI is calling UXA, he would transmit the following:

UXA UXA UXA DE TVI TVI TVI

The reply from UXA would then be:

TVI TVI TVI DE UXA UXA UXA

In the single-station call procedure, only one call sign, usually that of the called station, is used. For example, if ZYH is calling ARW, he would send ARW ARW ARW; when ARW answers, he would reply in the same manner, ARW ARW ARW.

Sometimes one particular call sign is assigned to a link, i. e., for intercommunication between two specific stations. For example, referring to Fig. 1, when Station 1 wishes to make contact with Station 10, he would send the link call sign BND repeated several times, and Station 10 would reply with the call sign BND.

In addition to the foregoing types of calls, there may also be used a collective call sign for calling several specific stations in a net; when such a call sign is used for alerting all of the stations in the net, it is called a *net call sign*. For example, Station 1 uses the net call sign RZK for reaching his three outstations, and Station 3 uses VJM as his net call sign.

In all of the foregoing procedures, split-call working might be employed As an example, we note in Fig 1 that Station 3 uses the call sign TVI when communicating with its superior, Station 1, or with Stations 2 and 4; however, when Station 3 is communicating with its own outstations, it uses the call sign ZYH.

Stations in a net are assigned one or more frequencies for radio communication; the allocation of frequencies is predicated upon transmitter characteristics, distance requirements, the time of transmission, and other factors. In *simplex* working, stations operate on a common frequency; in *complex* working, more than one frequency is used. In complex sending, stations are assigned *transmitting* frequencies, and each station uses its assigned frequencies to make contact with other stations; in complex receiving, stations are assigned *receiving* frequencies, and stations sending to a particular station use the frequency assigned to it.

The time of communication is an important factor in radio operations. Schedules for communication are established for those stations which pass comparatively little traffic, or which have an insufficient number of operators for free communication with all necessary stations; in such cases, schedules are arranged so that each operator may take care of several circuits at different times. Such schedules also permit maximum use of one frequency, without interference or confusion. When no schedules are in force, stations are free to contact each other at any time, either by setting the time for the next contact at the last transmission, or by maintaining a watch on assigned frequencies.

RADIO PROCEDURES

In radiotelegraphy the transmission of information is accomplished by means of Morse codes. In the case of countries whose alphabets

differ from the English alphabet, modifications of the international Morse symbols are introduced to take care of accented and other unique letters of the language.

Radio operators use certain signals and signs to facilitate operation and passing of traffic. The most common sets of operating signals, used in international practice, are "Q" and "Z" signals, which are three-letter combinations beginning with these letters. For example, QRU means "I have nothing for you," and QRU followed by a question mark (Morse $\overline{\text{IMI}}$) means "Do you have anything for me?" Besides these operating signals, various procedure signs are used by the operators, such as the following:

$\overline{\text{AR}}$	End of transmission	GR	Group count
$\overline{\text{AS}}$	Wait	$\overline{\text{IMI}}$	Repeat or question
$\overline{\text{BT}}$	Break	K	Invitation to transmit
C	Correct	WA	Word after
DE	From	$\overline{\text{VA}}$	End of schedule

In addition to the foregoing, radio operators may be provided with a specialized cryptosystem, usually in the form of a small chart (with row- and column coordinates) containing letters, digits, words, and useful short operators' messages.

In order to prevent enemy stations from entering a net and confusing its operations, authentication systems are used. In station authentication, challenges and replies are exchanged mutually by stations upon establishing initial contact; in message authentication, certain elements from the heading and from the message text are designated by prearrangement as test elements, and these test elements are validated by an authenticator symbol or symbols in the preamble.

In military communications, a single time designation is used to avoid the confusion that would result if each station used local time as reference. Normally, Greenwich Mean Time is used for all communications, although in some instances the time zone of the capital of a country is employed; in any case, it is usual practice to include the suffix letter of the time zone, as for example 231600Z meaning 1600 Greenwich Mean Time on the 23d of the month.

There are certain elements of the transmission which are standard for most radio operations. These are: (1) the call-up, or the procedural rules by which stations make contact with one another to prepare for the transmission of traffic; (2) the order of traffic, governed by rules which determine which station is to transmit its traffic first, and in what order; (3) the transmission of traffic, in a prescribed manner; (4) the receipting for traffic, in which the receiving station acknowledges receipt of messages; (5) corrections and services, to insure that the traffic transmitted and received is as garble-free as possible; and (6) the signing off, or the procedures prescribing the manner of terminating transmissions. Variations in the number and detail of the foregoing elements exist not only among various nations, but also among the military services of a particular country and among the different echelons of these services.

RADIO MESSAGES

Radio messages must carry pertinent information to insure proper handling in both the message center and the radio station. This information, almost invariably incorporated in the message externals, usually includes serial numbers of various kinds, date-time groups, precedence symbols, routing instructions, addresses and signatures, the group count, and other special instructions.

The number which is put on the message by the transmitting radio operator for reference purposes is known as the *station serial number* (abbr. NR); a number series may be assigned to all messages transmitted by a particular station, or separate number series assigned to messages passed on each communication link. *Message-center numbers* (abbr. MNR) are numbers assigned serially by a message center to all outgoing traffic, regardless of destination; these numbers are used for reference purposes between originating and receiving message centers. When messages are relayed, the station serial numbers change on each link of the communications path, whereas the message-center number usually remains constant. Other kinds of numbers are sometimes found in message externals, especially at the higher echelons, such as cipher-office numbers or radio-station in-desk numbers.

Precedence indicators or symbols for expediting traffic are either in the form of abbreviated plain text (such as "U" for Urgent) or in encrypted form as a group of letters or digits. Sometimes variants are provided for these indicators as a security measure, or these indicators may be subjected to encipherment.

When direct communication between two stations is not possible, routing instructions are usually incorporated in the externals of messages. Designations of locations or units in plain text may be utilized for this purpose, or call signs may be used for the routing, but, more usually, routing codes are employed which contain code groups for principal locations or units, as well as syllabary groups for encoding designations not in the body of the code. Similarly, when addresses and signatures are distinct from routing instructions, a separate scheme may be devised for the transmittal of this information, usually by means of codes.

PRELIMINARY NET RECONSTRUCTION

In the initial approach to a traffic analysis problem, traffic identified by the language of plaintext chatter, or by national characteristics of the transmission, as belonging to the target country is segregated into major homogeneous types on the basis of common operating characteristics, message formats, discriminants, chatter, or any collateral information. Thus traffic from army, navy, air force, and other nets may be isolated into distinct groups.

A preliminary grouping of stations is diagrammed from observed contacts between stations. Simultaneously, analysis is begun on the characteristics of the radio operations. As an example, let us assume that the groupings of stations in Figs. 2*a* and *b*, below, have been reconstructed from observed contacts on the transmitting frequencies in

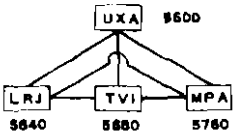

Fig. 2a.

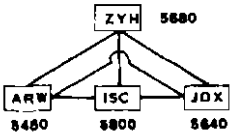

Fig. 2b.

kilocycles as indicated, and that we have made a mental note that, on the basis of procedural characteristics, UXA and ZYH are probably net control stations. We note that TVI and ZYH have the same frequency; if frequencies are assigned uniquely to target stations, then TVI and ZYH represent the same station. Tentative confirmation may be obtained if it is found that the serial numbers used by TVI interlock with those of ZYH, or if routing information on messages from TVI and ZYH shows identical originators; further confirmation may be obtained from chatter (wherein, for example, the operators at TVI and ZYH refer to the same person as their commanding officer), from direction-finding bearings, Morse-operator analysis, discriminant and indicator studies, etc. By continuing this method of analysis, we shall arrive at a portion of the diagram in Fig. 1, wherein TVI and ZYH are shown as split call signs belonging to one station. This example of approach is perhaps an oversimplification, but it is illustrative of the general methods followed.

ANALYSIS OF RADIO OPERATIONS

This phase of traffic analysis involving the study of the operating data and the elements of the transmission is, as previously stated, carried on concurrently with initial net reconstruction. When fragmentary nets have been put together, continuity over date-breaks is made possible by the analysis of radio operations.

Callsign analysis embraces the determination of the methods of generation, allocation, and rotation of call signs, together with the system of use. Call signs may merely consist of different random *n*-character groups, in which case no system of generation is recoverable, or they may be generated by a permutation table or similar scheme. The available call signs may be arranged in the form of a chart or in a book of tables, and stations may be allotted specified positions in the chart or book on, let us say, the first of the month; subsequent changes of call signs may be governed by following a prearranged route in the chart or book, or by the application of some mathematical formula. Callsign systems may also involve several sliding strips as a means of generation, with a convention prescribed for the manner of selection and rotation of the call signs derived from the strips. Regardless of the system of generation and rotation, when sufficient callsign continuity has been established, interpretation of the patterns and phenomena disclosed will permit recovery of the system.

Frequency analysis has the same general objectives as callsign analysis, *viz.*, the determination of the methods of selection, allocation, and rotation of frequencies, together with the system of use. When more than one frequency is assigned to each station, lower frequencies are generally used at night and higher frequencies during daylight, for technical reasons; certain of the frequencies may also be designated as standby frequencies. Frequency assignments may be published in chart form, with an initial allocation and rotation system similar to that used in callsign systems. Here again, continuity of frequencies will permit recovery of the system. Both in callsign and in frequency analysis, continuity may easily be obtained if some of the operating data or elements of the transmission change and some do not. Even if call signs and frequencies change daily, continuity may be established by taking into consideration any of the following: patterns of station serial numbers or message-center numbers; routing information; discriminants (especially one-time-pad discriminants which are usually unique for each link); procedural peculiarities (e.g., the use by a particular station of distinctive separator signs, tuning signals, etc.); chatter; schedules; service messages over a date-break; and direction finding and Morse operator analysis reports.

Procedure messages and chatter between operators are of particular interest in traffic analysis. When unknown procedure signals are used, or when procedure signals are encrypted, their meanings may be determined through observation and interpretation. As an elementary example, let us suppose that at 0915 an intercept operator hears TVI

send to UXA on 3800 kilocycles the procedural transmission XLC 1200, after which contact with TVI is lost, and that TVI is heard calling UXI again at 1158. The inference may be made that XLC means "I shall contact you again at ___ hours," followed by the time. Or again, let us suppose that after that same transmission, contact with TVI was lost, and that the intercept operator in searching for target stations on his receiver picks up TVI a few moments later on 4800 kilocycles. In this case, it may be inferred that XLC means "I am changing my frequency to ___ kcs," followed by a frequency designator which is to be multiplied by 4 to indicate the actual frequency.

The identification of preamble components is a relatively simple matter. If messages from Station A to Station B are sorted by intercept time, the station serial numbers should be in an ascending series (barring, of course, missed traffic), so that we look for such manifestations in elements of the preamble. If all the traffic sent from one call sign, regardless of direction, is sorted by file time (where this information is included in the preamble), the message-center numbers should be in an ascending sequence, with gaps caused either by missed traffic, or because the station concerned used more than one call sign, or because some messages may have been transmitted by means other than radio. The position of originator groups in the message preamble may be discovered by sorting traffic by transmitting station and noting the consistency of certain groups in a particular position; likewise, addressee groups may be identified by sorting traffic by receiving station and looking for a high rate of occurrence of some group or groups in a particular position in the preamble. The identification and interpretation of precedence indicators may be accomplished by studying a small volume of traffic emanating from one station and comparing the file times with the intercept times; when a series of messages are transmitted by a station one after the other, the messages with higher precedence are invariably transmitted first, and study of the traffic will give clues as to the meanings of these indicators. Sometimes preambles also contain groups indicating the security classification of the messages; these groups are often difficult to identify and interpret, but nevertheless a study of chatter and of the discriminants used on the various cryptonets will permit a solution.

As may be observed from the foregoing discussion, identification and partial solution of the elements of the preamble proceed simultaneously; further study and analysis will make possible a complete solution of these elements. Additional information on radio operations can be derived through study of schedules, textual features of encrypted traffic, cryptonets, and discriminants and indicator usage. Collateral information will be of assistance in these studies, as will information derived from cryptanalysis and other communication intelligence sources.

TRAFFIC INTELLIGENCE

The last phase of traffic analysis is the reconstruction of the complete enemy network in the form of an integrated diagram showing call signs, frequencies, and other technical data such as serial-number allocations, discriminants, etc. Identifications of unit organizations and their geographical locations are shown, which, when coupled with intelligence from all sources, will portray the enemy Order of Battle.

When changes in net structure take place, these may be brought about by the appearance of new units in a command or the deactivation or redeployment of old units. Changes in contact relationship may be indicative of impending moves; significant changes in traffic volumes or in cryptographic systems may be indicative of preparations for military activity.

CONCLUDING REMARKS

Traffic analysis furnishes much information on communications features of assistance in cryptanalysis, such as information concerning the originators and addressees of the messages, isologs and resends which result from cryptographic error, messages with potential crib value, and chatter pertaining to cryptographic matters.

Some traffic analysis items of particular interest to the cryptanalyst are the following:

When the group count is constantly checked by the enemy operators, this is usually indicative that the cryptosystem includes transposition as one of its steps.

When the date or file time is invariably checked, it is indicative that these elements are factors in key selection.

When a group in a particular position of the text or of the preamble is checked frequently, this may indicate that it is involved in key selection.

Rapid sending, with no requests for services by the receiving operator, is an earmark of practice or dummy traffic.

The general principles of traffic analysis have been presented briefly in the preceding paragraphs; however, as with cryptanalysis, a real understanding of these principles and techniques can come only with practical application.