（续）

法语	18	8	7	6	5	4	3	2	<1	
	E	AN	RSIF	UO	L	D	CMP	VB	F-Y	
葡萄牙语	14	13	12	8	6	5	4	3	2	<1
	A	E	O	RS	IN	DMT	UCL	P	QV	（空）
德语	18	11	8	7	5	4	3	2	<1	
	E	N	I	RS	ADTU	GHO	LBM	CW	（空）	
加泰罗尼亚语	15	12	8	7	5	4	3	1	<1	
	E	A	S	ILRNT	OC	DU	MP	BVQGF	（空）	
匈牙利语	16	13	8	6	5	4	3	<2		
	E	A	T	OS	LNZ	KIM	RGU	（空）		
意大利语	13	12	11	9	7	6	5	3	2	
	E	A	I	O	L	NRT	SC	DMOU	VG	
荷兰语	20	10	7	6	5	4	3	2	<1	
	E	N	IAT	O	DL	S	GKH	UVWBJMPZ	（空）	
西班牙语	13	9	8	7	5	4	3	1	<1	
	EA	O	S	RNI	DL	CTU	MP	GYB	（空）	

Kullback给出了如下8种语言的长为N的密文中单表双字母表。对于给定的语言，可由概率数据推导出固定值（表3-6）。

表3-6 单表和双字母报文

单表报文	二字母报文	三字母报文
英语	0.0661N(N-1)	0.0069N(N-1)
法语	0.0778N(N-1)	0.0093N(N-1)
德语	0.0762N(N-1)	0.0112N(N-1)
意大利语	0.0738N(N-1)	0.0081N(N-1)
H语	0.0819N(N-1)	0.0116N(N-1)
葡萄牙语	0.0791N(N-1)	
俄语	0.0529N(N-1)	0.0058N(N-1)
西班牙语	0.0775N(N-1)	0.0093N(N-1)
	随机报文	
0.038N(N-1)	0.0015N(N-1)	0.000057N(N-1)

3.13 德国的约简密码——流量分析

密码分析的一个姐妹科学就是流量分析科学。流量分析是信号情报分析的分支，信号情报分析研究信号通信的外部特征。

我们把信息用于：(1) 有效截获；(2) 帮助密码分析；(3) 在不了解特定报文内容的情况下判断情报的价值和级别；(4) 改进通信网的安全性。二次世界大战期间，破解德国密码的主要原因就是流量分析（当然，同样的原理也适用于美国和英国的密码）。德国军队有大量文件和无线电报文要按规定的方式传送给不同的军事单位。

3.13.1 组成要素

考虑到语言、通信程序符号和信号的区别，军用无线电通信系统包含六个标准的组件，它们是：(1) 呼叫；(2) 通信的顺序；(3) 通信的传输；(4) 通信的接收；(5) 纠错和服务；(6) 停止通信。

为了保证战场上和报务中心对报文的正确处理，某些信息往往以明文形式或用简单的代码发送。线路和日常的信息通常出现在报文的开头和结尾。这主要包括：(1) 顺序号、报务中心编号；(2) 组数；(3) 文电日期、时间（类似PGP签名）；(4) 线路系统-发方、收方和转发（提供明显标识参考位置）；(5) 优先级别（重要的信息需要标识，因此FLASH报文则表示紧急报文）；(6) 传输和交付程序；(7) 地址和签名；(8) 特殊说明。作为一条规则，德国的高层通信包含这些大部分要素，而低级别的通信可能把这些要素简化至最少。

我们可以从德国人处理报文顺序号的方式看出他们喜欢组织。任何由师一级发向战场上士兵的无线电报文都附有一个明文或矩阵密码的参考顺序号，这个序号可能是由报文作者、指挥部报务中心、信号中心或编码室、发报员或操作员加上去的。线路系统通常由代码和表示位置或单位的音节符号表组成。

我们看到在如今的电子邮件或文字处理系统中，通过可移植的桌面，这些信息变得更加容易处理，但是仍然可以运用流量分析。将上述的六个通信要素与现代计算机网络的数据包进行比较，我们按照OSI模型来考虑一下信息流，注意到所有六个要素在数据包头和用于路由信息的协议中都有其对应的东西。

美国密码学家擅长从德国人的密码网络中确定他们的战斗命令。流量分析不仅给出了战场上各个单位的位置，而且给出了各单位之间或单位集群之间的通信关系。在代码和密码的组成中，某些德国战斗命令允许使用纬度。这被证明是德国密码安全中的一个可以利用的错误。

3.13.2 用于密码分析

通过Crib报文、同底报文（Isologs）和字符，流量分析可以产生许多有用信息。*Crib*报文就是通过对外部特性的识别而获得了明文的部分信息。在德军各单位之间来回传送的命令“sitreps”（形势报告）对于美国的密码分析人员来说是最容易不过的了。明文形式的发送源、序号范围、密码网ID、报告类型、文件日期和时间、报文长度以及错误报文，彻底暴露了德军指挥命令。德军的战斗命令、部队部署和部队活动均由流量分析获得了。

如果相同的明文是用两种不同的系统加密的，就会出现同底报文的情况。造成这种情况的原因可能有多种，如接收方要求重复发送，或者预定的报文要发送给多个接收方，或者是译电员出现失误。美国的密码研究人员非常善于利用这种方法获取情报。

概括起来讲，流量分析就是寻找单位之间的联系关系，跟踪他们的活动，构造密码网，利用他们结构中的非随机性等等。美国的情报人员在第二次世界大战中非常成功地利用了这种方法来破解德国和日本的密码。

3.13.3 ADFGVX

“Weh dem der leugy und Klartext funkt”（那些撒谎并且以明文发送无线电报文的人应该受

到诅咒）——德国第五军团耶格（Jaeger）中尉。耶格是一位德国密码专家，他于1918年被派往法国以加强德国的密码纪律。讽刺的是，在1918年就是他名字中的两个“e”使美军的流量分析专家修复了代码变更。

ADFGVX是密码学历史上一个最有名的战地密码。最初是一个只有5个字母ADFGX的5 × 5矩阵，1918年1月1日对其进行了扩充，加入了第6个字母V。之所以选择这6个字母是因为它们在摩尔斯电码中的简洁性：A.-，D-..，F..-.，G--.，V...-和X-..-。弗雷德曼描述了关于1918年5月到8月间法国马恩河和兰斯战事活动的一个流量分析图表。这个分析就只基于ADFGVX密码中的流量。该密码仅限于德国师指挥部和陆军军团之间以及他们内部的高级指挥通信。

ADFGVX密码被认为是安全的，因为它在一个加密系统中结合了好的代替（双向分段或两部分分段加密系统）和好的换位。在该密码8个月的使用寿命期间，盟军只恢复出10个密钥（在10天的大流量内），并且解读了这些天内50%的报文。这些报文的解译影响了Ludendorff领导的德国15个师的推进，当时他们在巴黎以北大约50英里的Montdidier和Compiegne。著名的法国密码分析家Georges Painvin上尉的破解方法也只适用于两种特殊情况。盟军没有找到通用的破解该密码的方法，但在1933年弗雷德曼和SIS找到一个通用的破解方法。法国的吉维耶热将军也公布了一种适用于一般情况的破解方法。

Painvin破解的一份6月3日的报文改变了第一次世界大战的进程：发自德国Remaugies高级指挥部：Munitionierung beschleunigen Punkt Soweit nicht eingesehen auch bei Tag（英文就是：Rush Munitions Stop Even by day if not seen）。

密文的开头为： CHI-126:FGAXA XAXFF FAFFA AVDFA GAXFX FAAAG

这就等于在计划德军下一次的主要进攻之前告诉了盟军炮击的时间和地点。

3.13.4　对ADFGVX密码的分析

根据弗雷德曼的介绍，在当时只有三种不同的方法来破解这种密码。第一种方法需要两份或多份有相同明文开头的报文，以恢复换位。在第二种方法中，则需要两份或多份有相同明文结尾的报文，以破解密文代替部分的均匀分布的保护（分布越均匀，密文的随机性就越好，因此就越加难以破解密码）。德国人喜欢使用固定格式的措词，这一点在德军通信中也非常流行，因此可以在每天的通信中找到有类似开头和结尾的报文，或者二者兼而有之。第三种方法需要有完全相同长度的多份报文。Painvin在1918年4月破解五个字母版的ADFGX密码时就使用了前两种方法。

为免于我们低估这个密码的难度，我们仍可以借助于Painvin已做过的工作。3月21日下午4时30分，6 000杆枪同时向盟军的索姆河防线开火。5个小时以后，德军的62个师就已推进到前线40英里处。无线电通信量此时大增。Painvin仅仅截获了不多的几份ADFGX密报，较长的被分成了三部分以防止猜字法攻击。五个字母，可能是一个西洋跳棋盘？也许吧！是简单的单表代替密码吗？不是，分布太均匀了。

德国报文的奇怪之处在于其第一部分均有相同的报文碎片，且顺序相同。Painvin认为这很可能是由使用相同的密钥对报文开头进行换位所致，或者是由换位表各列相同的开头所致。Painvin将密报分成如下组：

chi-110: (1) ADXDA (2) XGFXG (3) DAXXGX (4) GDADEF

chi-114: (1) ADXDD (2) XGFFD (3) DAXAGD (4) GDGXD

他总共使用了20个分组密文来进行分析，以恢复换位密钥。利用长列置左的原则，他发现第3、6、14和18密文片断在左边，平衡的串在右边。再利用其他有相同结尾的报文，他将这样的列放于左边。这样分析正确吗？好像不对！随后他又利用其他18份截取的报文以并置60个字母：AA、AD等等，通过频率统计，他发现了一个单表代替，而且他还发现第5列和第8列是互逆的。

Painvin构造了一个棋盘的大概结构，并假设字母的顺序。

	A	D	F	G	X
A					
D			e		
F					
G					
X					

因为报文是20个字母，顺序应当是从侧面开始由上而下反复，这意味着在加密过程中边坐标落在了第1、3、5的位置，所以他利用频率统计特性将它们进行了分离。在经过48个小时的艰苦工作之后，Painvin终于将明密文字母配对成功并还原了加密表（棋盘），破解了当时世界上最难的战地密码：用分段来保护密码，明文字母被分成相同大小的分组，必然分散其普通特征。换位进一步以特定的方式分开了这些特征，从而使通常帮助恢复换位的线索不明显。

3.14　阿拉伯人对密码学的贡献

Ibrahim A.Al-Kadi博士在1990年就关于阿拉伯人对密码学的贡献的问题向瑞典皇家技术学院提交了一篇优秀论文。

Al-Kadi博士称阿拉伯科学家Abu Yusuf Yaqub ibn Is-haq ibn为Sabbah ibn 'omran ibn Ismail Al-Kindi，此人在大约公元750年写了一本关于密码学的书，叫《Risalahfi Istikhraj al-Mu'amma》（密码报文破解手稿）。Al-Kindi在书中介绍了密码分析技术、密码的分类、阿拉伯语音学和词法，最重要的是描述他使用几种统计技术进行密码分析的方法（此书大概要比其他密码学方面的书籍早300年，它也要比帕斯卡和费马早800年提出概率和统计的概念）。

Al-Kadi博士还报告了Al-Khwarizmi（780—847）的数学著作，此人引入了一些通常的技术术语，比如零、密码、算法、代数和阿拉伯数。十进制数系统和零的概念最初是由印度人提出来的。

在9世纪早期，阿拉伯人就将婆罗门的佛教经文由梵语译成阿拉伯语。新的计数系统很快就被从中国到西班牙乃至整个伊斯兰帝国所接受。Al-Khwarizmi关于算术的书由切斯特（Chester）的Robert、哈利法克斯（Halifax）的John和意大利比萨的Leonardo进行翻译，译书中强烈提倡使用阿拉伯计数系统，以取代先前标准的罗马计数系统（I，V，X，C，D，M）。

罗马计数系统让人觉得很麻烦，因为它没有零（空）的概念。零的概念，我们认为很自然的事情，但在中世纪的欧洲却是有反对意见的。在梵语中，零被称为*sunya*或“空”。阿拉伯人

将印度人的零译成了阿拉伯语中的sifr。欧洲人采用了这个概念和符号，但没有采用其名字，而是将其变换成了拉丁语中的*cifra*和*cephirium*。在意大利语中零被称为*zefiro*、*zefro*和*zevero*。*zevero*简写为*zero*。

法语中新造了一个词*chiffre*，并采用了意大利语的*zero*。英语使用*zero*和*cipher*，是因为ciphering可看成一种计算的方法。德语中采用单词*ziffer*和*chiffer*。

zero、sifr和cipher的概念对于普通的欧洲人来讲过于模糊和容易混淆，以至于人们在争论中会说“讲清楚点，别用cipher”。*Cipher*的本意就是隐藏明文消息，简单说就是加密。Al-Kadi博士断言，阿拉伯语的*sifr*发展成了欧洲人“加密”的技术术语。

3.15　Nihilist代替

由于某种原因，俄罗斯不允许囚犯在牢房里使用计算器。俄罗斯囚犯相互之间也被禁止交流。为了瞒骗监狱管理人员，他们发明了一种“敲击”系统，以表示简单的棋盘上的行和列（Polybius方阵在英语中是5 × 5方阵，俄语有35个字母，则使用6 × 6方阵）。比如:

	1	2	3	4	5
1	U	N	Ij	T	E
2	D	S	A	O	F
3	M	R	C	B	G
4	H	K	L	P	Q
5	V	W	X	Y	Z

密钥字＝United States of America

密钥字中的重复字母在方阵中只使用一次，即首次出现时才出现在方阵中。

明文: g o t a c i g a r e t t e

密文: 35 24 14 23 33 13 35 23 32 15 14 14 15

囚犯们记住了正确的数，每分钟大约能“说”10到15个单词。这种方法的一个好处就是它可以借助好多种媒介来达到通信目的，任何可以指示数字的方法都可使用。

密文字母就由写在一起的字母的数量指示，间隔用手写的空格，或者是书写时向上的笔划、向下的笔划、拇指指甲印，所有微小的东西都用来在监狱内外传递秘密。这种系统在刑事机构是普遍的。美国的战俘（POW）在越南还使用它。密钥字的换位还会提供混合的加密表:

B L A C K S M I T H

D E F G N O P Q R U

V W X Y Z

按列的顺序读取则是:

B L V L E W A F C G Y K K Z S O M P I Q T R H U

所以Polybius方阵就变成了:

	1	2	3	4	5
1	B	D	V	L	E
2	W	A	F	X	C