Mandatory Escrow Schemes, Law and Politics

by John Kasdan

Columbia Institute for Tele-Information
Graduate School of Business
809 Uris Hall
New York, New York 10027
(212) 854 4222

Mandatory Escrow Schemes, Law and Politics.

John Kasdan *

The concept of key escrowing first became national news at an extraordinary event, a product announcement by the White House. On April 16, 1993 a press release describing the Clipper chip was released by the Office of the White House press secretary. The Clipper chip was a tamper-proof chip, produced by the Mykotronx Corporation, incorporating a classified algorithm, SKIPJACK, and a so called unit key which would be split into two "shares" so that law enforcement personnel, with appropriate authorization, could obtain the two parts which, when combined, would enable them to execute wiretaps on transmissions encoded by the chip. The distinctive feature of this scheme was that the holders of the two pieces of information, the so-called escrow agents, could, individually, not deduce the unit key from the information they posessed. Although there had been previous academic discussion of the escrowing concept, mostly based on the work of Silvio Micali, the Clipper chip was the first device incorporating the concept which was widely displayed.

The original proposal for the Clipper Chip called for only two escrow agents, the National Institute of Standards and Technology and a division of the Treasury Department, both governmental agencies. The procedure for producing the shares suffered from the disadvantage that both shares were produced at the same point, in a secure facility so that it was unlikely that neutral observers would be able to verify that the two shares were kept separate. To further add to distrust of the chip, the SKIPJACK algorithm was pointedly not made generally available for evaluation, although a small group of cryptographers were shown it. Thus some people feared that there might be a "trap door" in the algorithm which would make it possible for government officials to break Clipper transmissions, even without going through the request permissions which were described as being necessary to unite the shares.

The multiple problems of the Clipper chip proposal led to many expressions of distrust of the system by various segments of the cryptographic community. RSA Data Security, the premier firm in non-governmental encryption even deemed the issue worthy of a "Sink the Clipper" T-Shirt. Eventually, Matt Blaze showed that there was a flaw in the protocol used by Clipper

* Lecturer in Law, Columbia University School of Law. kasdan@cunix.cc.columbia.edu. I would like to thank Richard Pierce, Gerald Neuman, Gerard Lynch and Richard Uviller for many helpful discussions.

and its brethren which could defeat the methods by which taps were supposed to be carried out. Although various governmental sources connected with the Clipper chip proposal denied that the protocol failure had serious consequences for Clipper's use in its primary intended application, telephony, the Clipper seemed rather quickly to die, presumably of embarassment, after the Blaze paper was published in the middle of 1994.

Shortly thereafter, Vice President Gore wrote a letter to Rep. Maria Cantwell which was widely interpretted as abandoning the Clipper chip initiative. Although Gore has since suggested that his letter was misinterpretted, one part of it was clear at the time, and apparently remains the policy of this administration. That is that the administration wants any general use of strong encryption to include key escrowing for the purpose of aiding law enforcement.

Although the Clipper chip proposal was presented as voluntary for any use except for communications with the Government, the logic presented for the escrowing scheme would seem necessarily to require that key escrowing be mandatory. As the April 13 press release said,

> "The chip is an important step in addressing the problem of encryption's dual-edge sword: encryption helps to protect the privacy of individuals and industry, but it also can shield criminals and terrorists. We need the "Clipper Chip" and other approaches that can both provide law-abiding citizens with access to the encryption they need and prevent criminals from using it to hide their illegal activities."

However, if Clipper, or more generally key escrowing, were to be merely voluntary, it would seem likely that criminals would make use of non escrowed strong encryption so as to be able further to hide their communications from law enforecment agencies.

The Clipper chip was, clearly, a public relation fiasco. Although no part of the project was handled really well, probably the worst aspect of the project was the attempt to establish a proprietary encryption algorithm as a standard, if not actually as the unique, mandated encryption method. If there is any single guiding tenet of modern cryptography, it is that security should lie only in the key and not in an attempt to keep the encryption algorithm secret. Too much of the history of cryptography, such as the breaking of the German ENIGMA during World War II, recounts "impossible" occurences whereby the enemy has gained possession of a "secret" code machine and thereby compromised an entire set of messages. Many people no longer believe that anything is gained by hiding algorithms. Thus SKIPJACK was, properly, distrusted from the moment it was announced.

Furthermore the method of escrowing which was suggested for the Clipper proposal involved user's keys being in the hands of two Federal agencies. With incidents like Watergate and the rifling of State Department passport records for political purposes all too frequent in our recent history, it is not surprising that people felt that shares escrowed under the Clipper proposal might not be very safe and might even be compromised at their creation.

The weakness of the Clipper proposal, moreover, contrasted with the richness of concurrent developments in the non-governmental cryptographic area. The increased power of such chips as the Power PC, the Alpha and the Pentium greatly speeded up the computationally intense public key encryption methods. Triple DES and IDEA, along with the surprising revelation of the RC4 algorithm, followed by the release of the RC5 algorithm gave users a wide choice of symmetric data encrypting techniques. With all these choices available, it is not surprising that the multiply flawed Clipper proposal was rejected, especially when the logic of the proposal strongly suggested that Clipper might move from a voluntary option to the only available encryption method.

Escrowing is a rich technique. Clipper is a system where a single classified encryption method comes with keys established in a government controlled inaccessible facility, which are then divided into a small number of shares held entirely by the government. It is possible to have an escrowing system with none of those characteristics. That is, there could be key escrowing with heterogeneous publicly disclosed algorithms where users established their own keys and shares, which were then escrowed with arbitrary numbers of agents, at least some of whom might be non governmental organizations.

As an example of escrowing, consider escrowing the private key in the first public key encryption system invented, Diffie-Hellman key passing. The solution to the problem of passing a private key through a public communications channel depends on there being no efficient way to solve a certain mathematical problem. In 1978, Diffie and Hellman found an appropriate problem which could be applied to the key passing problem. Let p be a large, perhaps 150 digit, prime number. That is, a number divisible only by 1 and by itself, like 3 or 71. Efficient ways to find such numbers are known. We may assume that p is known to everyone who is interested. Let b be some number greater than 0 but less than p, of approximately the same size as p, i.e. also a 149 or 150 digit number. Assume that b is also generally known. Finally, let r be another large number less than p. But assume that r is not known. The difficult problem is: given p, b, and $b^r (mod p)$ (which means raise b to the r-th power and then subtract as many multiples of p as necessary to make the result greater than 0 and less than p. I.e.,$x = b^r (mod p)$ is the unique number,$0 \leq x < p$, such that there is a whole number n with $x + np = b^r$.)

By analogy with normal logarithms, this is called the discrete log problem. The reason that it is easy to find regular logarithms is that for $x > 1, log x < x$. But that is not true for discrete logs. For example, $41^{16}(mod 59)$ is 3, and 3 is less than 16. As an example, $41^{r}(mod 59) = 49$; what is r? (The answer is 15, but it is hard to find except by trial and error.)

Given the difficulty of solving the discrete log problem, here is how key passing works. Assume p and b, as above are generally known. The first party tells the second that he wishes to communicate, privately, with her. She agrees. He then secretly chooses a large number r, which is less than p-1. He then tells her $b^{r}(mod p)$. She, secretly, chooses a large s, less than p-1, and communicates $b^{s}(mod p)$. Now he can raise her number to his power, reduce $mod p$, and get $(b^{s})^{r}(mod p)$, while she can raise his number to her power and reduce, getting $(b^{r})^{s}(mod p)$. But both of these numbers are equal to $b^{rs}(mod p)$, so both parties know the same number and the first few digits of these equal numbers can be used as their private key for a DES transmission. Up until the present, no easy way to solve that problem has been found, and many competent mathematicians have tried.

It is next necessary to explain how the private key, the exponent r, can be escrowed. The idea, due to Silvio Micali, is to have some number, $n > 1$, of escrow agents, each of which will hold some piece of information (the "share") such that if all of the agents combine their shares it will be possible to reconstruct the exponent r, but if any proper subset of the agents uses only the shares they collectively have, not only will they not be able to reconstruct the key; they will not even have any advantage in solving the discrete log problem over anyone who did not have access to their shares.

In the case of Diffie-Hellman, this is extremely easy to do. Remember that the public key associated with the private key, r, is $b^{r}(mod p)$. Now since, by Fermat's little Theorem, b^{p-1} is congruent to 1, $(mod p), r$ need only be known $(mod (p-1))$. The escrowing scheme, then, is to choose n-1 numbers, a_i, randomly from the interval $[1, .., n-1]$ and choose a_n such that $\sum a_i$ is congruent to $r, (mod (p-1))$. The share of the i-th escrow agent is a_i. The proof that this fits our requirement for an escrowing scheme (that no proper subset of the escrow agents can reconstruct the private information from their shares is very similar to the proof that a one time pad cannot be broken: for any possible private key, r, and any set of shares, $a_i, 1 \leq i < m$, there exists another set of shares, $a_{m+1}, ..., a_n$ such that $\sum a_i$ congruent to $r(mod (p-1))$.

In other words, just as with one time pads, for any cyphertext there exists some possible plaintext and some possible pad which together yield the cyphertext, for any proper subset of shares there is a complementary set of shares which yields any possible private key. So in the case of one time pads all that one can know is the (maximum possible) length of the plaintext and with the escrow system described one learns nothing about the private key.

Furthermore, escrow schemes can be designed which enable some number less than all of the escrow agents to reconstruct the private key. Such a scheme could, for example, be based on on the fact that n+1 points on the graph of an n-th order polynomial of one variable completely determine the polynomial, while for any n given points in the plane with different x coordinates, one may find some n-th order polynomial passing through any other point with an x coordinate different from those of the n given points. Thus if the share given to each of $M > n + 1$ agents was a pair, $(x_i, f(x_i)), x_i \neq 0$, and the private key was $f(0)$, than any n+1 agents could, by use of the Newton interpolation formula, construct the private key, while any n, or fewer, agents would derive no aid in finding the private key from their shares. Obviously this scheme can be generalized to have different classes of agents so that, for example, any 2 from class 1 and 4 from class 2 (but no fewer from either class) could reconstruct the key.

Similar, though in detail more complicated, escrowing schemes exist for RSA encryption and for other public key systems. In fact, the above described systems, working $(mod\, 2^{56})$, could be used to escrow keys for the public key DES method.

The reason I have described escrowing in such detail is to support the proposition that escrowing is as theoretically secure as the encryption system whose keys are being escrowed. The next question, then, is whether an actual escrowing scheme would be secure.

The advantage of multiple agents is simply that a conspiracy among agents to release keys would become more difficult when there would have to be a greater number of partcipants for a successful conspiracy. Legal incentives, in the form of criminal penalties for unauthorized disclosures and, perhaps, rewards for agents revealing attempts to obtain shares could also increase the security of shares. Private escrow agents might also be perceived as having common values with their users and would therefor be seen as less likely to compromise shares. Somewhat fancifully, one can imagine a world with Chabad, Shabazz, Aryan Nation and other special-interest escrow agents.

Such an escrowing regime, of course, introduces an extra vulnerability to a user of encryption. A potential eavesdropper could attempt to obtain the shares of as many of the agents as were required to reveal the key, but doing one theft from the user would probably be vastly easier than pulling off a multiplicity of thefts from organizations which existed solely for the purpose of safeguarding data. There would, of course, be incentives for employees of escrow agents to discover keys. Presumably many of the keys would make possible such lucrative activities as instituting money transfers from a user's account to that of the faithless employee. However similar incentives exist today for, say, bank employees and a combination of legal sanctions and internal controls seems to hold down the number of embezzlements. And, in the escrow situation, there is the additional complication that the embezzler would also have to coordinate with insiders at a large number of other escrow agents.

Law enforcement officials, attempting to listen in on transmissions for which they have not obtained authorization, would be in the position of any other intruder.

In other words, a broadly diversified escrow system with users creating and delivering their own shares to agents of their own choice would be clearly more secure than the world in which most of us now live; that is, a world in which we regularly make unencrypted transmissions that can be intercepted by anyone with the technical ability to do so.

Thus there seems to be some justification for the claim, often made by the FBI and others in favor of key escrowing schemes that all that such a scheme would do is to leave law enforcement in the position it is in today, capable of executing a warrant, but still facing the stringent requirements of obtaining a warrant under Title III. (Similar claims were made in behalf of the recently passed Digital Telephony bill, 47 U.S.C. §1001 et. seq.) If this claim is true, and mandatory escrow would merely preserve the status quo ante, what reasons could there be for opposing such a proposal?

Several possible answers to this question come to mind. The first is that it is only the existence of private strong encryption that makes it possible for an individual personally to defeat law enforcement wiretapping. Until the decision in Katz v. U.S., 389 U.S. 347 (1967), wiretapping was not even considered a search and did not require a warrant. Starting from that baseline, the legislative requirements of Title III, requiring not only a showing of probable cause, but also that the wiretap is either necessary for an investigation or that all other methods are too dangerous, was a vast increase in protection for individual privacy. And judicial and legislative protection was, until the advent of personal encryption, the only type of protection that electronic privacy could expect. At the present, though, the potential for an individual to deploy cryptography which might take a major nation years to break, makes self help possible.

This technological breakthrough just happens to come at a time when many, according to the polls most, Americans have reached a new low in their estimation of the government. Accordingly, it is not surprising that many people would react angrily to any restriction on their ability to shield their affairs from agents of the government.

However, I wonder whether that reaction is not the property primarily of the cryptographic elite. Cryptography has not yet become a major national issue. Crime however has. I am by no means certain what the outcome would be if a sizable portion of the American people understood the debate over key escrowing, but it would not surprise me if the issue was seen as one of protection against criminals, and if there was overwhelming support for restricting encryption to whatever degree law enforcement felt was necessary.

Discussions of possible Constitutional infirmities to mandatory escrowing laws, if they are not undertaken purely as intellectual excercises, presuppose that such a law could be passed. Ruth Bader Ginsberg said about the right to abortions that it would have been better if such a right had been established through the political system, rather than through the Courts. I think the same is true of encryption. Further, since the Constitutional challenges to a mandatory escrow law appear to be far less than slam dunks, it might be advisable for those who would oppose such a law to start considering how a political opposition could be made.

The first step in such a strategy is clearly to identify the costs imposed by mandatory escrowing. One such cost is the symbolism inherent in saying that the government can compel individuals to organize their lives for the purpose of expediting wiretaps. Although it can be argued that the laws regarding the removal of vehicle identification numbers (VIN's), 18 U.S.C. §511(a)(1), and firearm serial numbers, 26 U.S.C. §5861(g), have already established the principle, the symbolism of a law affecting all telephones would be vastly more powerful. Of a lesser degree is the cost in efficiency that a mandatory escrow law would impose. Presumably such a law would require re-escrowing upon changing one's key. This would presumably be no harder than having a new key certified in a key certification scheme, a field in which much work has been done. But for those who were willing to depend on more casual key transmittal, the requirement of escrowing would impose some cost. It seems likely, however, that the costs of the Digital Telephony bill in actual costs and possible deterioration of service will be far greater than any costs associated with key escrowing, and it must be noticed that the digital telephony costs have not been a major political issue.

The other possible attack on a mandatory escrow act which occurs to me is to focus on what benefits law enforcement expects to get from such a bill.

At a recent meeting of the Association of the Bar of the City of New York, Special Agent Jim Kallstrom of the New York Office of the FBI said that the FBI was

"in the information business of collecting criminal information. Now who are these crimi-

> nals? You all know who they are, you read the papers. They're the people who blew up the World Trade Center, the cost of which is estimated by the Port Authority of New York as somewhere around 5 Billion dollars to the GNP of this area. They're the terrorists. They're the drug dealers. And the kidnappers, the extortionists. They're the child pornographers. You know they're all these people who we're chartered to put out of business for the good of all of us. For the good of society. One way we do this is with electronic surveillance. A larger and larger part of the information that goes on the table in these crimes is obtained through electronic surveillance. So the notion that we're giong to allow criminality to be beyond the scope of the law, beyond the scope of a search warrant; the notion that they're going to have a reservoir immune from the judicial process, the idea that we're going to give up on the notion of collecting the critical information about the kidnapping of a young child that might save that child's life. We're not yet ready to say that the genie's out of the bottle. ... The notion that technology is just some sort of snake that's slithering along, and that it should go wherever it likes, wherever the brain trust is going to let that go, without a realization that there's public policy issues here – public safety issues here – is, I think, a very naive approach."

Transcript of January 19, 1995 Meeting, at 8-9.

The intersting thing about this statement, besides its somewhat overheated tone, is that it identifies only one group with any specificity, the World Trade Center bombers. And that is a case in which it is clear that the current ability to conduct electronic surveillance without any technical impediments did the FBI no good at all. Had the bombers not been concerned about collecting the $150 deposit on the truck they blew up, all of them might have escaped the country, just as the mastermind of the group did. In short, the best defense against a possible mandatory escrow law might not be Constitutional, but might rather consist of a political argument that the gains from such a law would not equal the costs it would impose. I suspect that such an argument would require convincing people that the symbolic disadvantage of limiting encryption would be greater than the help that such a law would give law enforcement, and I suspect that such an argument will prove hard to make, since crime is a hot-button item and, as of yet, encryption is not.

Having made the point that a political attempt to forestall the passage of a mandatory escrowing bill might be a better strategy than a legal challenge to such a bill after it was passed, it must still be noted that there are interesting Constitutional questions that such a bill would raise.

I. First Amendment Problems

Suppose that Congress passed a law saying that any person who used encryption over any electronic communications lines had to have an escrowing scheme for that method approved by the NSA (escrow schemes for several popular methods like Diffie-Hellman and RSA would be approved in advance of the law taking effect; for other systems the user would have to propose the scheme and have it approved) and escrow her key with any number, up to some maximum, of registered escrow agents before using the encryption method. If the user changed her key, she would have to escrow the new key before she could use it.

The law would provide that, upon obtaining a warrant in accordance with the current provisions of Title III, law enforcement agents could obtain the shares from the escrow agents of the subject of the warrant and carry out their wiretap. The law would provide substantial penalties if, after obtaining the warrant, the agents determined that the subject was not using her escrowed key for transmitting.

The law would also contain requirements for escrow agents and would provide for penalties if the agents either refused to honor warrants for shares, or notified clients that a request for shares had been made. [1]

The first, and crucial point, to notice about my proposed law is that it does not attempt to regulate the content of any speech. That is, the law applies to any message send by telephone or e-mail, regardless of the content of that message. If the message is to be encrypted, it must be encrypted by a system for which there exists an approved key escrowing scheme, and the key must have been appropriately escrowed.

The Supreme Court has recently stressed that the question of whether a law regulates speech on the basis of the content of that speech is crucial to determine the level of scrutiny which will be given to that law. "[T]he First Amendment, subject only to narrow and well-understood exceptions, does not countenence governmental control over the content of messages expressed by private individuals. Our precedents thus apply the most exacting scrutiny to regulations that suppress, disadvantage, or impose differential burdens upon speech because of its content. ... In contrast, regulations that are unrelated to the content of speech are subject to an intermediate level of scrutiny."

Turner Broadcasting System v. F.C.C., 114 S.Ct. 2445, 58-59 (1994).

The Court, in trying to determine whether the regulations in question (which require most

[1] In fact, there already exists one regulation restricting the use of cryptography. 47 C.F.R. §95.412(a) states that "You may use your Civilian Band station to transmit two way plain language communications. Two way plain language communications are communications without codes or coded messages."

cable operators to carry the programming of local broadcast television stations) were content based or content neutral noted that "the rules benefit all full powered broadcasters ... be they commercial or non-commercial, independant or network-affiliated, English or Spanish language, religious or secular." id. at 2460.

The Court also stressed that the regulation was "based only upon the manner in which speakers transmit their messages to viewers, and not on the messages they carry." id.

In the case of the proposed encryption statute, it is again only "the manner in which speakers transmit their messages" which is being regulated. The law is not intended to apply only to political speech or anti-feminist speech or any other particular type of message which might be transmitted. Instead it applies to any message which is transmitted in a certain way, namely encrypted.

The appropriate test for a content-neutral rgulation of speech was set forth in United States v. O'Brien, 391 U.S. 367 (1968). Such a regulation will be sustained if, "it furthers an important or substantial governmental interest; if the governmental interest is unrelated to the suppression of free expression; and if the incidental restriction on alleged First Amendment freedoms is no greater than is essential to the furtherance of that interest." 391 U.S. at 377. The Turner Court endorsed, and in fact directly applied, this test. 114 S.Ct. at 2469.

It is to be noted that, formally, this test is not a "balancing" test. It is more deferential to government interests than that. It does not demand that the governmental interest in question "outweigh" the speech interest that the regulation impairs. It merely requires that the interest be "important or substantial" and that the method adopted to meet that objective interferes with speech interest no more than is necessary to achieve the end.

At this point it becomes necessary to consider the important or substantial purpose which the government might claim was the justification for a key escrowing bill. The statement of Agent Kallstrom, quoted above, is an example of governmental articulation of purpose. Despite the facial inconsistency of using the World Trade Center case as a justification for greater abilities to wiretap, given the deferential quality of the O'Brien test, it seems to me quite likely that a court would find the asserted need for the government to be able to bypass strong encryption to be an important or substantial interest. And, if the legitimacy of the interest is granted, it is hard to see how speech interests could be affected any less than they would be by a distributed escrow system. In fact, it may be a more interesting question to ask whether, if the concept of key escrowing had not been developed, a total ban on the non-governmental use of encryption could be justified for the purpose of allowing law enforcement officials to be able to tap conversations.

The basic argument I have presented is that a key escrowing requirement would not be considered to be content related, and that the government could articulate important reasons for having such a law. I have then asserted that, in general, a deferential standard is applied in such conditions, which would lead to such a law being upheld in the face of a first amendment challenge. However, it is certainly true that there have been cases where a law that did not on its face appear to restrict speech on the basis of content has, nonetheless, been found unconstitutional. One such area, arguably relevant to cryptography, has been anonymity.

[Discussion of Talley, NAACP, etc. Discussion of languages, using Meyer and 9th Cir. Arizona Official Language case.]

[Discussion of 4th Am. and administrative searches and exigent circumstances.]