The Birth and Death of the Orange Book

Steven B. Lipner

Over the past 50 years, US government computer security strategy has shifted focus from government-funded research and system development to evaluation of commercial products. By tracing the history of the Trusted Computer System Evaluation Criteria (TCSEC) or Orange Book during this period, this article covers the role of government agencies, vendors, and policymakers in determining IT system security requirements and development.

This article traces the origins of computer security research and the path that led from a focus on government-funded research and system development to a focus on the evaluation of commercial products. That path resulted in the creation of the Trusted Computer System Evaluation Criteria (TCSEC), or Orange Book. The TCSEC evaluation regime resulted in a great deal of investment on the part of both the US government and commercial vendors, but in the end, the TCSEC and the underlying security model were largely abandoned.

The history of the Orange Book provides a cautionary tale that is relevant today to technologists and policymakers who seek to mandate improved cybersecurity. This article draws on public sources, including technical reports and conference proceedings that report government and vendor perspectives, as well as the proceedings of the online forum shared by vendors and TCSEC evaluators. Although many of these sources reflect the public positions of the US government, the article does not reflect internal government deliberations or documentation.

Beginnings

By the late 1960s, government agencies, like other computer users, had gone far in the transition from batch processing to multiuser and time-sharing systems. The US Department of Defense (DoD) Advanced Research Projects Agency (ARPA) was a primary funder of research into time-sharing,[1–3] and ARPA program managers made the power of the new models of computing evident throughout the government. The growing availability of commercial products that were capable of supporting multiple batch-processing job streams, time-sharing users, or transaction processing terminals (or more than one of the three) facilitated the transition to multiuser systems. By 1970, DoD was planning a major procurement of mainframe computers referred to as the Worldwide Military Command Control Systems (WWMCCS) to support military command operations.

Defense and intelligence agencies were among the early government adopters of the new models of computing, and they were quickly faced with the need to use the new systems to process classified information. This was a new world. The agencies had operated batch-processing systems in locked vaults, but the transition to multiuser systems posed both basic challenges—physical security of terminals, protecting communications, and controlling access by remote users—and more advanced challenges that included keeping simultaneous users apart and protecting classified information.

The desire to meet the more advanced challenges emerged early. The Air Force's Military Airlift Command (MAC), for example, provided the military services with a largely unclassified air cargo and passenger service but on rare occasions was required to classify some of its missions using the same aircraft and crews—for example, in cases of military contingencies or special operations. By 1970, MAC had articulated a requirement to process classified information on its soon-to-arrive WWMCCS mainframes while

 Published by the IEEE Computer Society 1058-6180/15/$31.00

allowing users without security clearance to access classified information (uncleared users) access to the mainframes.[4]

The national security community responded to the challenges in two ways: the Office of the Secretary of Defense commissioned a study of the policy and technical issues associated with securing computer systems, while ARPA funded the development of a prototype secure operating system that could process and protect classified information.

The study effort was organized as the Defense Science Board (DSB) Task Force on Computer Security under the chairmanship of the late Willis Ware. Its membership included technologists from the government and defense contractors as well as security officials from the DoD and intelligence community. The task force met between 1967 and 1969 and produced a classified report that was made available to organizations with appropriate security clearance beginning in 1970.[5]

The Ware Report, as the DSB task force report came to be called, provided guidance on the development and operation of multiuser computer systems that would be used to process classified information. Viewed from the perspective of 2014, the Ware Report does not strike the reader as wildly outdated. Part of the report focuses on policy and anticipates the need for system certification measures (approval to operate) that are not radically different from current practices as represented by the recently issued DoD Instruction 8500.[6]

The technical sections of the Ware Report focus on the difficulties of building highly secure computer systems. The report points the way for future research by directing attention to operating system ("supervisor") security, referring to the promise of the then-new Multics operating system, and exploring the importance of integrating labels for classified information into the operating system. Systems that could manage information at multiple classification levels and users at multiple clearance levels were referred to as "multilevel secure" systems. Two of the report's key findings had especially significant impact on future efforts:

- A (multilevel) secure open system [one that can adequately protect classified information from uncleared users] cannot be provided by contemporary (late 1960s) technology.
- Since commercially designed supervisors and operating systems have not included security control, it is to be expected that the average commercial software will not provide the … capabilities required.[7]

The Ware Report was originally released as a classified (confidential) document, so its circulation was limited to those in the national security community. However, it had a significant impact by, in effect, telling the defense organizations such as the MAC that they could not have the open secure systems that they required.

In parallel with operation of the DSB Task Force, the System Development Corporation (SDC) was moving forward on the development of the ARPA-funded ADEPT-50 prototype time-sharing system. It is important to view ADEPT-50 in the perspective of the late 1960s. Mainstream commercial operating systems supported batch processing and sometimes supported multiprogrammed batch-processing and transaction processing. Time-sharing was still a relatively new model of computing, and ARPA had funded several projects aimed at constructing time-shared computer systems.[1–3]

ADEPT-50 was SDC's second time-sharing system (the first was simply referred to as the time-sharing system for the IBM military AN/FSQ-32 computer). The key features of ADEPT-50 included operating system features intended to protected labeled classified information.[8] Several defense programs considered using ADEPT-50 operationally, and SDC sought to continue its development. But an Air Force-sponsored review in 1969[9] found that the risks and limitations associated with ADEPT-50 were too great and that the Air Force would be better served by exploring other alternatives—perhaps based on other research projects, perhaps on commercial products. After the review was completed, several projects that had planned to make use of ADEPT-50 project wound down and the ADEPT-50 project at SDC eventually came to an end.[8] The demise of ADEPT-50 was significant because it deprived the Air Force of a potential solution to its multilevel security requirements and because it foreshadowed the move to commercial products and away from government-developed operating systems.

The Air Force

The Ware Report sent the clear message that the nation's best computer security experts advised against attempting to operate open (classified information, some uncleared users) multilevel secure computer systems.

But Air Force requirements for multilevel systems remained outstanding. The MAC requirement for an open multilevel system was the most demanding, but it was not the only requirement. The Air Force Data Service Center (AFDSC) in the Pentagon had articulated a requirement for a multilevel time-sharing system that would allow programmers with Secret clearances to develop models that would support defense analysts who would apply the models to Top Secret data.

In the early 1970s, Air Force requirements for the development of new computer system capabilities were addressed to the Air Force Electronic Systems Division (ESD) in Massachusetts. ESD relied in large part on the technical advice and support of the MITRE Corporation, a federally funded research and development center (FFRDC). Thus, the MAC and AFDSC requirements found their way in late 1970 to MITRE's Information Systems Department. An early MITRE report[4] suggested alternative approaches to meeting the MAC requirement without developing a new multilevel secure operating system in hopes that these approaches might avoid the problems the Ware Report characterized as intractable. It quickly became evident to the ESD/MITRE team, however, that the problem of multilevel security was even more difficult than initially realized and that significant research and development would be required to satisfy multilevel security requirements.

Although ESD and MITRE had a fundamentally conservative outlook on problems of technology, as evidenced by the recommendation that the Air Force wind down its work with ADEPT-50,[9] the Air Force requirement for multilevel security remained open, and the users at MAC and at AFDSC showed no signs of forgetting it. Thus ESD and MITRE began a series of projects aimed at exploring the feasibility of developing multilevel secure systems. The first such project—executed by MITRE—sought to build a multilevel system with limited function, a "secure communications processor."[10]

ESD also augmented its in-house staff focused on security during 1971. Roger Schell, who was an Air Force major and had recently earned a PhD in computer science from the Massachusetts Institute of Technology, joined ESD in the summer of 1971 and was able to sponsor a number of projects. Schell's dissertation research had been conducted as part of the Multics development group at MIT,[11] and he sought to determine whether Multics—a research system commercially supported by Honeywell and with security as a basic design goal—could contribute to solving the multilevel security problem.

Schell's commanding officer at ESD, an Air Force colonel named Edmund Gaines, pressed Schell and his support contractors at MITRE to get "expert opinion" behind them. This pressure resulted in the creation of a panel of experts led by Jim Anderson and Ted Glaser, both veterans of the DSB (Ware) Task Force.[12] The panel's members were drawn from government, industry, and academia with significant overlap with the DSB Task Force members and technical advisors. Members drawn from the National Security Agency (NSA) and the Defense Intelligence Agency (DIA) were recruited both to take advantage of their established expertise in security and to ensure that the panel's results would carry an implied endorsement by the intelligence community.

The Anderson Panel met six times between February and September 1972 and produced a report that outlined a comprehensive R&D program in computer security. The primary focus of the report was the development of computer systems that could be trusted to enforce multilevel security, although it also recommended research in areas including secure terminals and storage media protection. The report recommended a technical direction based on the concept of a reference monitor, a mechanism that would

- mediate all access to information based on a security policy,
- protect itself, and
- be small and simple enough to be subject to complete analysis and testing.

The Anderson Report[13] justified the investment in secure computer systems on the basis of a cost estimated at $100 million per year resulting from the absence of secure multilevel computer systems. The report outlined a research program (the report termed the elements of the program advanced and exploratory development) that would culminate in procurement specifications for secure computer systems. Read from the perspective of 40 years later, the report is surprisingly silent about the existence of a commercial IT industry that develops products for a broad market and that considers government requirements as only one input to its product planning decisions. The report presumes a program lifecycle common to government technology

efforts of its era: once the research is completed, the government will release a request for proposals for secure systems, and capable suppliers will apply the technology and respond with acceptably secure systems.

Even before the Anderson Report was released, the ESD team had started a series of projects with the aim of executing the recommended research program. Projects at MITRE[14] and Case Western Reserve University[15] sought to develop mathematical models of multilevel security. These projects formalized the notion of mandatory security in which subjects (active elements in a computer system such as users and processes) must possess security authorizations (in effect, clearances), objects (data repositories such as files) must possess security labels (in effect, classifications), and highly classified information is prevented from being communicated to uncleared users. Mandatory security is distinguished from discretionary security models in which individual users are free to share or disseminate information to which they have access.

Schell had proposed an implementation of the reference monitor concept that he called a security kernel,[12] and another project at MITRE sought to build a prototype for the Digital Equipment PDP-11/45 minicomputer.[16] Finally, Schell was able to convince AFDSC that Multics, although not a realization of the reference monitor concept, was a step along the way that could be used to meet the AFDSC need for limited multilevel secure applications (Top Secret data on a system that supported Secret cleared users). Thus, AFDSC procured a Multics system from Honeywell. The contract for AFDSC's Multics system included requirements for enhancements that would integrate the MITRE model of multilevel security and laid the groundwork for future operating systems that implemented multilevel security but not a security kernel.[17]

The full realization of the vision in the Anderson report was to be a version of Multics that incorporated a security kernel. Schell developed a complex set of arrangements involving funding from the Air Force, ARPA, and Honeywell and participation by Honeywell, MIT, and MITRE in a program known as Project Guardian. The aim of Project Guardian was to restructure Multics so that the core of the system and the component on which the system's security would depend was a security kernel that was an implementation of the Anderson Report's reference monitor concept. Honeywell would also build a separate secure minicomputer: the Secure Communications Processor (SCOMP). The SCOMP was to be a communications front end, with its own security kernel, that would replace the Honeywell Datanet 355 communications processor that implemented the interface between the Multics mainframe processor and users' terminals.

Schell's approach was controversial within the Air Force, where some senior executives believed that industry would solve any problem that was worth solving. Largely as a result of these views, the funding for Project Guardian was never fully supported by Air Force headquarters. By 1976, the funding arrangements had collapsed and Project Guardian had been cancelled. The work on the restructured Multics system ceased early in the design phase, and the kernel-based system was never completed. The legacy of Project Guardian lived on in the Orange Book, however, and it had a powerful impact on the national computer security strategy for the next 25 years.

Steve Walker and the Computer Security Initiative

ARPA had funded the development of Multics at MIT[3] and retained an interest in computer security into the 1970s (and beyond). In mid-1974, a computer scientist named Steve Walker moved from NSA, where he had worked on early computer networking projects, to become the ARPA program manager for the Arpanet and computer security research. Walker cooperated with Schell in funding Project Guardian and became convinced that the reference monitor and security kernel approach to security advocated by the Anderson Report was at least a plausible way to address the problem of multilevel security.

NSA had been interested in computer security since the 1960s. Hilda Faust and Dan Edwards from NSA's research organization served on the Anderson Report panel, and well before the completion of the Anderson Report, NSA had begun funding the Provably Secure Operating System (PSOS) project at SRI with the aim of producing a formally verified secure operating system.[18]

When Project Guardian was cancelled, Walker at ARPA stepped in to ensure the continuation of funding for security research. While NSA pursued its own somewhat distinct security research directions (PSOS and a series of projects focused on network security and applications of cryptography), the NSA

research staff also collaborated with Walker in the execution of ARPA's research projects.

By the late 1970s, ARPA was funding or cofunding at least four projects aimed at the development of secure or multilevel operating systems:

- KVM, an SDC project that sought to integrate multilevel security and a security kernel into the IBM VM/370 virtual machine monitor;
- KSOS, a Ford Aerospace project that sought to integrate multilevel security and a security kernel into Unix;
- SCOMP, a project at Honeywell that sought to complete the secure communications front-end hardware from Project Guardian and use it to host a multilevel secure security kernel-based operating system similar to Unix; and
- the University of California, Los Angeles (UCLA) Data Secure Unix and its predecessor UCLA secure virtual machine monitor, which was aimed at implementing security kernel based but not multilevel secure operating systems for the Digital Equipment PDP-11/45 minicomputer.

With the possible exception of the UCLA effort, each of these projects sought to build usable systems that could actually be deployed.[19] The developers, however, were all government research contractors rather than IT vendors. (Honeywell was a vendor, but the SCOMP project was executed by Honeywell's government contracting organization.) The research community had not yet tackled, or perhaps fully realized, the challenge of "technology transfer" to commercial products.

The late 1970s saw a significant shift both in the government's approach to computer security and in Walker's role. In 1978 Walker moved from his research program management role at ARPA to a staff position at the Office of the Secretary of Defense (OSD). Much of his work at OSD was concerned with DoD communications and networking strategy, but he retained responsibilities related to DoD computer security policy and strategy.

At some point in the late 1970s, Walker drove a fundamental shift in DoD strategy for acquiring secure systems. Although previous research and prototype efforts had been conducted by government research contractors and focused either on building systems that could be deployed or on developing specifications that could be met by government contractors,[13] the new strategy focused on defining evaluation criteria that commercial off-the-shelf (COTS) IT vendors would meet as they built products.[20]

The origin of the shift in strategy is not completely clear. Walker's oral history does not refer to the change in direction.[21] George Jelen's history cites a letter from the late James Croke, a MITRE vice president, that mentions the shift.[22] Sheila Brand[23] attributes the change in Walker's views to his participation in a 1977 workshop on auditing of computer security, although the workshop report[24] does not explicitly call for the evaluation of commercial products.

It is clear that the new strategy was more popular—or at least less unpopular—with COTS computer vendors. Schell's oral history describes IBM's reaction to the notion of building a security kernel.[12] Government security evaluations would still constrain the kinds of products that vendors could sell, but they would presumably not be as intrusive as a requirement to build systems that conformed to government specifications.

Walker's and the DoD's change in strategy were reflected in the creation of the DoD Computer Security Initiative, which sought to define criteria and process for evaluating the security of commercial IT products and to create an infrastructure for conducting evaluations. As part of defining the criteria and process and bringing COTS vendors on board with the concept of evaluation, the Computer Security Initiative sponsored informal technical exchanges between DoD and FFRDC security experts and vendor secure system development teams. Many major vendors of the era participated in these exchanges.[25]

The initiative established a series of meetings that began in July 1979 as the Seminar on the DoD Computer Security Initiative Program[26] and, by 1985, evolved into the National Computer Security Conference.[27] These meetings served to bring together the government and MITRE program managers who were creating the evaluation scheme and the COTS vendors whose products would be subject to evaluation, along with research contractors and academics who were continuing to participate in the multilevel secure prototype efforts and develop research tools that could be used to verify product security. The conferences continued as gatherings for a growing computer security community through 2000.[28]

Creating the Orange Book

The Computer Security Initiative was Walker's way of building both a technical foundation and a bureaucratic base for the strategy of evaluating COTS products. Walker considered establishing an evaluation center at the National Institute of Standards and Technology (NIST) rather than NSA. In the end, the DoD management decided that the DoD Computer Security Center (later the National Computer Security Center and hereafter referred to as the NCSC) should be created as part of NSA. The bureaucratic maneuverings that led to the creation of the NCSC at NSA (and not at NIST) are best documented in Jelen's report.[29] The center was established in 1981 with Air Force Colonel Roger Schell, leader of the early 1970s work at ESD, as its deputy director.

Work had been going on well before the establishment of the NCSC toward the definition of evaluation criteria that could be used to assess the security of COTS products. Much of this work was conducted, or at least reported, by the computer security team at MITRE, which was still active and still populated with many veterans of Project Guardian and its predecessors. MITRE's Grace Nibaldi authored a MITRE report in 1979 that laid out the initial plans for the evaluation of COTS operating systems.[30]

The Nibaldi paper, and the MITRE presentations at the first seminar on the Computer Security Initiative, make it clear that multilevel security and the ideas from the early Air Force projects had not been forgotten, and in fact, they had a significant impact on the emerging evaluation criteria. The Nibaldi paper places great emphasis on the importance of mandatory security. Like the Orange Book to follow, it defines seven levels of evaluated products with the lowest, least-secure level (0) reserved for "unevaluated." In the Nibaldi scheme, all but level 1 (the lowest level that actually undergoes evaluation) must include features for extensive mandatory security.[30]

The creation of the Orange Book was a major project spanning the period from Nibaldi's 1979 MITRE report to the official release of the Orange Book in 1983. Draft evaluation criteria may have been created between 1979 and 1982, but if so they do not appear to have been circulated widely. The first public draft of the evaluation criteria was the Blue Book released in May 1982.[31] The structure and identification of evaluation classes (levels in Nibaldi's terms) in the Blue Book are almost the same as those that finally emerged in the Orange Book, but they are considerably different from those in the Nibaldi report. From least to most secure, they are as follows:

- D: Minimal Protection
- C1: Discretionary Security Protection
- C2: Controlled Access Protection
- B1: Labeled Security Protection
- B2: Structured Protection
- B3: Security Domains
- A1: Verified Design
- A2: Verified Implementation

The inclusion of classes C1 and C2 indicates recognition on the part of the NCSC that most existing commercial products did not incorporate mandatory access controls suitable for multilevel security applications. NCSC believed that vendors with existing products would seek to have them evaluated at class C2 but then progress to higher classes over time as the market grew and product developments were completed.

Class A2 (Verified Implementation) would have required formal verification of an evaluated product down to the source code level as well as other rigorous measures including development by a team of cleared personnel. It appears in each of the three draft evaluation criteria[31–33] but is transformed to a speculative section titled "beyond A1" in the official Orange Book ultimately released.

One feature of the Blue Book draft appears significant with the benefit of hindsight: the description of each of the evaluation classes includes a subsection for examples that describes systems—real, under development, or proposed—that would be expected to meet the evaluation requirements of that class. Roger Schell's address to the May 1982 Seminar on the Computer Security Initiative[34] similarly included real-world examples of candidates for each class:

- D: Unevaluated products.
- C1: Most "mature commercial operating systems" with specific reference to Unix.
- C2: The then-commercially available RACF, ACF2, and Top Secret add-on products that added security controls to the IBM MVS operating system.
- B1: The Blue Book refers only to retrofitting mandatory access controls to a mature operating system, but Schell's address specified the version of Honeywell's GCOS 3 operating system that was

modified to meet the requirements of the (early 1970s) DoD WWMCCS procurement for COTS mainframes.
- B2: Multics as modified for use at AFDSC.
- B3: The version of Multics that would have been produced had Project Guardian been carried to completion.
- A1: KVM, KSOS, SCOMP, or the communications processing systems developed for the Air Force SACDIN program.

The examples were removed from the internal NCSC draft[33] and do not appear in the Green Book final draft[32] that preceded the release of the Orange Book. As the next section will show, the absence of specific examples of systems that met the requirements of each class forced (or enabled) the NCSC evaluation teams to interpret the Orange Book without reference to real-world examples.

One significant issue that arose during the development of the Orange Book concerned the specific requirements for mandatory security policies—at what class would they be introduced and how rigorously would they be enforced. The language of the Nibaldi report, although brief, suggests the introduction of fairly rigorous mandatory policies at level 2: there is a reference to "flow controls" that would restrict the potential for an untrusted program to compromise classified information in violation of policy.[35] The definition of class B1 in the Blue Book and the internal draft, in contrast, requires security labels but specifically excludes a requirement for enforcement of mandatory policy that would prevent an untrusted program from compromising classified information by writing it to an object readable by a lower cleared user or process.[36,37]

The final resolution on "B1 mandatory security," as reflected in the Green Book draft and the Orange Book, reinstated the requirement (from the Nibaldi report[30]) that the evaluated product prevent untrusted programs from writing classified information to objects readable by lower cleared users or processes.[38] The rationale for the change, apparently not documented at the time, was to provide a consistent programming model for applications that had to operate in environments where security labels were present. The assumption at the time was that use of mandatory security would be pervasive, and it would make no sense to require application developers to accommodate labels in one way if an application was running on a class B1 system and in another way on a class B2 or higher system.

As the Orange Book was being completed, class B1 was referred to informally as "mandatory access control training wheels" by NCSC staff and their advisors. NCSC believed that mandatory controls were the right answer for secure computer systems, vendors would rush to build systems that incorporated mandatory controls, and customers would buy them. The assumption behind "training wheels" was that class B1 would serve as an entry-level introduction to mandatory controls and that vendors would build B1 systems and then move on to build systems at higher evaluation classes to meet the presumed demand. The NCSC mind-set, as illustrated by the Schell's 1982 address and by the programs for the seminars and subsequent National Computer Security Conferences, was that the "action" and the interesting evaluations were at evaluation classes B2 and above.

Evaluations, Interpretations, and Surprises

The Orange book was published in August 1983. Sheila Brand was the primary author[39] and several people inside and outside of NCSC were core contributors to its development. These included Grace Nibaldi (Hammonds) and Peter Tasker of MITRE; Dan Edwards, Roger Schell, and Marvin Schaeffer of NCSC; and Ted Lee of Sperry Univac. A number of people from government, government contractors, and vendors, including Jim Anderson, Steve Walker, Clark Weissman, and I were cited as reviewers who influenced the content of the final product.

With the official release of the Orange Book, the NCSC was "open for business" and vendors that had been cooperating with the Computer Security Initiative since the late 1970s began the process of preparing and submitting products for evaluation. Most vendors employed individuals or small staffs who were dedicated to product security and evaluation. Some of those individuals are named in the acknowledgements section of the Orange Book,[40] and many were frequent speakers at the seminars on the Computer Security Initiative and the National Computer Security Conferences.[41] Many had participated in the informal technical exchanges with government and FFRDC personnel previously mentioned here.

Given their exposure to, and in many cases influence on, the concepts underlying the Orange Book, and given the examples cited

in the Blue Book draft and Schell's address, these individuals believed they understood what it would take for their employers' products to complete evaluation successfully. They soon began to encounter surprises.

The first surprise for some of the vendors was coincident with the first announcement of evaluated products. IBM's RACF product, cited in the Blue Book as an example of a candidate for class C2, completed its formal evaluation in 1984 with an evaluation in class C1.[42] Available documentation does not specify the reasons for this outcome, but discussions at the time centered on RACF failing to include controls over "object reuse" (assurance that a newly created file would not contain data from previously deleted files). IBM engineers believed that the performance impact of this control would be excessive, although the competing CA ACF2 product did include controls over object reuse and completed a C2 evaluation at roughly the same time as RACF.[42]

RACF's problem with object reuse should have been evident from a reading of the Orange Book, but less obvious problems were soon to follow. One class of problem had to do with the language of the Orange Book. The drafts were written, and to a large extent reviewed, by individuals who had been working on computer security since the late 1970s or before. As such, they had a common language and common set of understandings based on experience with the Air Force and ARPA security programs and informal vendor exchanges. That language was reflected in the Orange Book; for example, the requirements for class C2 refer to discretionary access control over "named objects (e.g. files and programs)" and for B1 to mandatory control over "all subjects and storage objects under its [viz. the operating system's] control (e.g. processes, files, segments, devices)."[43]

The language regarding named objects and storage objects presumably made sense to its authors, but the authors were not the ones applying the criteria. When the NCSC went into full operation, evaluations were the responsibility of teams of government and FFRDC evaluators who reviewed vendors' products and decided whether they met the requirements of a particular class in the Orange Book. Differences of opinion between evaluators and vendor staff quickly arose, especially in cases where vendors were attempting to make existing products meet the requirements of class C2 and B1. A few examples illustrate the situation:

- Was it allowable for a class C2 or B1 system to include objects that were not subject to access controls and auditing? If so, which objects?
- Were interprocess communication channels "named objects" and subject to the C2 and B1 requirements for access control and auditing?
- How and to what extent should security controls such as login and auditing be applied to computer operators who worked in machine rooms and had physical access to mainframe systems, printed output, and storage media?

The NCSC had procured a Multics timesharing system to support its team of evaluators and to provide a secure shared computing and communications environment for evaluators, vendors, and consultants. Multics provided a "forum" facility comparable to a present-day threaded discussion facility. In 1985, the NCSC evaluation staff created a "Criteria_Discussion" forum to support open discussions of the Orange book and its meaning among evaluators and vendors. The forum accumulated more than 2,500 entries between 1985 and 2000 on a range of topics dealing with evaluations at all Orange Book classes. (The examples in the bulleted list earlier were taken from the forum.[44])

Many of the forum discussions show a clear difference in perspective between evaluators (who were trying to apply the requirements of the Orange Book in a rigorous and consistent manner) and vendors (who were trying to build and ship products to meet the evaluation requirements). One quote from the forum (entry 794 from March 1988) illustrates the evaluator perspective: "'Security relevant' refers to anything which the evaluation community feels may be relevant to system security."

Most of the forum entries are technical in nature, and it would be incorrect to say that the tone is strictly "evaluators versus vendors." But an adversarial tone is evident in some of the transactions from the mid- and late 1980s as the vendors come to realize that the process of developing evaluated systems would not be as simple as they'd believed.

Another example of the changing character of evaluations concerns documentation. The Orange Book documentation requirements for classes C2 and B1 are minimal:

- a security feature user's guide,
- a trusted facility manual (administrator's guide),

- test documentation (for testing of the system's security features), and
- design documentation (limited to a "philosophy of protection," descriptions of interfaces between modules of the evaluated product, and (at B1) an informal model of the system's security policy and an explanation to show that the system's protection mechanisms satisfy the model).

The documentation requirements at higher classes of the Orange Book are more extensive, as might be expected. But by the mid-1980s, evaluators were insisting on "design documentation" even for class C2 and B1 systems, making the argument that without such documentation, they could not understand the systems well enough to evaluate how they met the Orange Book requirements to their satisfaction. COTS vendors' documentation practices were (and are) often uneven, and product components might be uncovered by documentation, or the documentation might not be updated to represent the product "as built." From the vendors' perspective, extensive documentation should not be required for systems with a moderate or commercial-grade security objective. The evaluators had the final say, however, so vendors added staff or hired contractors to produce design documentation that was used to satisfy the evaluators rather than to guide developers in their implementation of products.

A final example comes from a former vendor employee who reported more obscure instances of debates with the evaluation teams that resulted from the unusual nature of his employer's system. The system in question was a distributed system in which the trusted computing base (TCB) was a dedicated microcomputer that ran only evaluated software:

- Was a capability-like discretionary access control mechanism that met the Orange Book requirements for fine-grained access control in an unusual way satisfactory to meet the requirements for class C2? The evaluators required the vendor to build a tool that reported access rights in a way that they found satisfactory—or at least comprehensible.
- Did a computer that was entirely dedicated to administrative functions (this evaluation was conducted after PCs became available) require an isolated TCB? Even though no untrusted code or users could attack the system or its software, the evaluator position was "yes."

The former vendor employee's perspective is that if the system under evaluation did not meet the evaluators' understanding of how a product should work, the evaluators "made up" an interpretation of the Orange Book. The NCSC eventually formalized the interpretations process, but vendors found it opaque and one-sided.[45]

C2 by '92

When the Orange Book was released in 1983, its developers' expectation was that vendors would quickly get their products evaluated at class C2 and rapidly modify their C2 products to add the mandatory security features needed to meet class B1. Meanwhile, they would begin the "real work" of building the systems at classes B2 through A1 that were needed to achieve multilevel security. Although not all vendors were believers in this progression, many were. A review of the papers in the National Computer Security Conference proceedings of the 1980s reveals multiple hints at development work on systems at class B2 and beyond.

By 1990, the picture was much different. Only Multics and SCOMP had completed evaluations as general-purpose operating systems at class B2 or above. (A secure LAN product had also completed evaluation at class B2.) Three products had been evaluated at class B1 (one had completed two evaluations on consecutive versions) and 12 C2 evaluations had been completed (again, some products had completed multiple evaluations). Table 1 summarizes the history of Orange Book evaluations from 1984 to 1998.[42]

The limited population of evaluated products at class B2 and above was largely attributable to limited demand for such systems by government (or other) customers. The last of the high-security evaluations by a major vendor (Digital Equipment) was cancelled in 1990 because of an insufficient market.[46,47] IBM built a class B2 version of the Xenix operating system for the PC in the mid-1980s and then gave it to Trusted Information Systems (a security consulting company founded by Steve Walker) rather than continue to market it.[21,48] The history by Donald MacKenzie and Garrel Pottinger presents a good picture of high-security systems (especially class A1) and of the challenges that confronted organizations that sought to build, evaluate, and market A1 systems.[47]

Table 1. Orange Book evaluations, 1984–1998.

	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
C1	1														
C2	1	1	3	2	3	1	1		2	1	6	3	2	5	1
B1						2	2	3	2	3	7	5	2	1	1
B2		1					1		1	1	1	1			
B3									1		1	2			1
A1	1							1			2				
Total	3	2	3	2	3	3	4	4	6	5	17	11	4	6	3

Perhaps as a result of the paucity of high-class evaluated products or in response to the small number of evaluated products overall, the late 1980s saw a shift in focus by the NCSC with a new slogan: "C2 by '92." The objective was to create government demand for C2 systems and thus encourage vendors to populate the Evaluated Products List with such low-security systems. The formal policy underlying "C2 by '92" required agencies that processed national security information to use class C2 systems. The policy was issued by the National Telecommunications and Information Systems Security Committee in 1987.[49]

Table 1 suggests that the "C2 by '92" initiative worked; although there are some anomalies in the data (a few database and network component evaluations as well as multiple evaluated versions of several products), the numbers of products that completed evaluations grew markedly. Because most vendors evaluated their standard operating system products at class C2, the demand side of "C2 by '92" worked by default: as a practical matter, it was hard for a government customer to acquire a commercial operating system that had not undergone a C2 evaluation.

Evaluations at class B1 also increased (Table 1), although repeated evaluations make this trend appear more significant than it actually is. One assumption that had driven the direction of the Orange Book was that the broad population of commercial customers would require and value the mandatory security controls that the Orange Book required. Although some briefings were presented and a few papers written,[50] by 1990 it had become evident that commercial customers did not require and would not use mandatory security controls.

Informal reports suggest that mandatory controls were useful in the AFDSC scenario, but the Naval Research Laboratory's Military Message Experiment[51] demonstrated that a system based on the Bell-LaPadula model was almost unusable and challenging to use securely in a dynamic multilevel secure email scenario—exactly the sort of scenario that would become common in the emerging world of networked computers. Theoretical questions were also raised about the mathematical model itself: without exceptions to the model ("trusted processes"), systems that implemented the model were unusable, but the exceptions undermined the soundness of the model.[51,52] The realization and acceptance of the reality of these limitations by the leadership of the NCSC was probably a factor in the decision to shift emphasis to C2.

The change of emphasis to C2 systems coincided with the disappearance of products in classes B2 and above. Versions of SCOMP completed multiple evaluations in the 1990s, and Trusted Information Systems completed two class B2 evaluations of the Secure Xenix system, but there were no new starts on systems targeted at classes B2 or above.

The Impact of the Orange Book

If the objective of the Orange Book and the NCSC was to create a rich supply of high-assurance systems that incorporated mandatory security controls, it is hard to find that the result was anything but failure. As of 2014, it is still possible to buy an updated SCOMP (now labeled STOP-OS and sold by BAE systems), but that product is the only survivor (see www.commoncriteriaportal.org). Oracle's Trusted Solaris and Red Hat's SE Linux incorporate mandatory security controls and might be candidates for class B1,[53,54] but the usage of those products does not appear to be widespread—some specialized applications in the national security community.

If the objective of the Orange Book and NCSC was to raise the bar by motivating

vendors to include security controls in their products, the case for success is stronger. Most major vendors did seek C2 evaluations for their operating systems, and discretionary security controls as well as the other features that class C2 required are common today. (Of course, many of the vendors whose products completed class C2 evaluations no longer exist today—a result of the radical changes in the IT market through the 1990s.) It is an open question whether vendors would have introduced those features even without the NCSC or whether the product impacts of the specific interpretations of discretionary security imposed by the NCSC were worth the cost and effort.

By the late 1990s, the era of the Orange Book was drawing to a close. The NCSC evaluation process had only been open to US vendors, and in response, other countries had created their own domestic evaluation schemes that competed with the NCSC. Many countries outside the US had agreed to follow the international Information Technology Security Evaluation Criteria (ITSEC) with the result that vendors were required to evaluate twice—once with the NCSC and once in a European country. Discussions about an international evaluation scheme began in the early 1990s, and the International Common Criteria for Information Technology Security Evaluation (Common Criteria, or CC), which replaced both the Orange Book and ITSEC, was signed at the 1999 National Computer Security Conference. Today, 17 countries conduct evaluations under CC and an additional nine countries accept the results of CC evaluations (www.commoncriteriaportal.org).

Through the first few years of the CC, evaluations were similar to Orange Book evaluations, requiring extensive "design" documentation and frequent "interpretations" by evaluation agencies. The emergence in the late 1990s of an industry of vulnerability finders who demonstrated security problems with commercial products made it evident that evaluated products fared no better under attack than any others.[55] As early as 2005, some of the CC schemes began to question the effectiveness of the CC process and a search for alternatives began.

The CC members considered a number of alternatives and eventually came to the conclusion that there was no benefit derived from the detailed product examinations that had required extensive design documentation. Instead, the CC evaluation schemes have chosen to go down a path that narrowly specifies product security features and the tests that will verify their correct functioning.

In a further reversal of the direction of the Orange Book, in early 2014, the US CC scheme (the successor to the NCSC) released a statement[56] that the broad evaluation of operating systems—the process that was the heart of the Orange Book and NCSC—does not provide sufficient security benefit to justify its cost. Instead, evaluations will be limited to tightly specified functional testing of specific operating system features.

What then was the legacy of the Orange Book?

- The Orange Book set the precedent for the government evaluation of commercial IT products.
- The Orange Book undoubtedly raised vendors' awareness of security and of the government's interest in security.
- The Orange Book diverted significant effort on the part of the security research community, and some effort on the part of vendors, into blind alleys: mandatory security and formal models of mandatory security, excessive documentation, and discretionary security features that were never used by customers.
- The Orange Book set governments down a path that led, after 20 years, to international cooperation in product security evaluation.

On the whole, the negatives appear to outweigh the positives. Government did not represent a sufficient market to fundamentally change the way that vendors built products, and the specific changes the Orange Book required were not valued by commercial or government customers. Perhaps the long-term benefits of international product security evaluation—a clear legacy of the Orange Book—will shift the balance.

Acknowledgments

My biggest challenge in writing this article was tracking down the references that would confirm or, in more than a few cases, refute my memories of the days of the Orange Book. My personal library includes many contemporaneous papers and conference proceedings, but there were some significant gaps. Those gaps were closed through the generous assistance of Sheila Brand, the principal author of the

Orange Book, who allowed me to reproduce her copies of the Blue, interim, and Green drafts and provided valuable comments on the change in strategy from government development to evaluation of commercial products; Jim Donndelinger, an Orange Book evaluator who is still in the evaluations business and who tracked down and provided a copy of the Dockmaster Criteria_Discussion forum; and Jeremy Epstein, who shared experiences as a vendor leading products through Orange Book evaluation. Once the draft article was done, Fred Schneider and Mary Ellen Zurko were kind enough to review it and provide critical comments. Their comments helped me make my points more clearly and to defend them better. Responsibility for the content and any errors is of course mine. Finally, my thanks to the Charles Babbage Institute for organizing the workshop where my original paper was presented and to the National Science Foundation for sponsoring the institute's project on computer security history. The views expressed in this article are personal and do not reflect the position of my employer.

References

1. D.G. Bobrow, "TENEX, A Paged Time-Sharing System for the PDP-10," *Comm. ACM*, vol. 15, no. 3, 1972, pp. 135–143.
2. F.J. Corbató, "An Experimental Time-Sharing System," *Proc. AFIPS Conf., Spring Joint Computer Conf.*, 1962, pp. 335–344.
3. F.J. Corbató and V.A. Vyssotsky, "Introduction and Overview of the Multics System," *Proc. AFIPS Conf., Fall Joint Computer Conf.*, part 1, 1965, pp. 185–196.
4. S.B. Lipner, "MACIMS Security Configurations," MITRE, 6 Jan. 1971.
5. W.H. Ware, "Security Controls for Computer Systems," RAND Corp., Feb. 1970; www.rand.org/pubs/reports/R609-1/index2.html.
6. T.M. Takai, "Cybersecurity," Dept. of Defense, 14 Mar. 2014; www.dtic.mil/whs/directives/corres/pdf/850001_2014.pdf.
7. Ware, "Security Controls for Computer Systems," pp. 26, 30.
8. C. Weissman, "Security Controls in the ADEPT-50 Time-Sharing System," *Proc. AFIPS Conf., Fall Joint Computer Conf.*, 1969, pp. 119–133.
9. ESD-MITRE Adept Study Group, "ADEPT Study Report," MITRE, 1969.
10. P. Tasker and D. Bell, "Design and Certification Approach: Secure Communications Processor," MITRE, 1973.
11. R.R. Schell, "Dynamic Reconfiguration in a Modular Computer System," MIT, June 1971; http://publications.csail.mit.edu/lcs/pubs/pdf/MIT-LCS-TR-086.pdf.
12. R.R. Schell, "Oral History Interview by Jeffrey R. Yost," Univ. of Minnesota, Charles Babbage Inst., 1 May 2012; http://conservancy.umn.edu/handle/11299/133439.
13. J.P. Anderson, "Computer Security Technology Planning Study," Deputy for Command and Management Systems, HQ Electronic Systems Division (AFSC), 1972.
14. D. Bell and L. LaPadula, "Secure Computer Systems: Mathematical Foundation," MITRE, 1973.
15. K.G. Walter et al., "Initial Structured Specifications for an Uncompromisable Computer Security System," Case Western Reserve Univ., 1975.
16. W. Schiller, "Design of a Security Kernel for the PDP-11/45," MITRE, 1973.
17. J. Whitmore et al., "Design for Multics Security Enhancements," Honeywell Information Systems, 1973.
18. P. Neumann et al., "A Provably Secure Operating System," Stanford Research Inst., 13 June 1975; http://csrc.nist.gov/publications/history/neum75.pdf.
19. Dept. of Defense Computer Security Initiative, "Seminar on the DoD Computer Security Initiative Program," Nat'l Bureau of Standards, 1979, pp. G-1, H-1, I-1.
20. DoD Computer Security Initiative, "Seminar on the DoD Computer Security Initiative Program," 1979, pp. C-1.
21. S.T. Walker, "Oral History Interview by Jeffrey R. Yost," Univ. of Minnesota, Charles Babbage Inst., 8 Nov. 2012; http://conservancy.umn.edu/handle/11299/144021.
22. G.F. Jelen, "Information Security: An Elusive Goal," Center for Information Policy Research, Harvard Univ., 1985, pp. II-77; www.pirp.harvard.edu/publications/pdf-blurb3d04.html?id=238.
23. S. Brand, email comm. with S.B. Lipner, 29 May 2014.
24. Z.G. Ruthberg, "Audit and Evaluation of Computer Security," Nat'l Bureau of Standards, 1977.
25. P.S. Tasker, "Trusted Computer Systems," *Proc. IEEE Symp. Security and Privacy*, 1981, pp. 99–100.
26. DoD Computer Security Initiative, "Seminar on the DoD Computer Security Initiative Program," 1979.
27. Nat'l Bureau of Standards and Nat'l Security Agency, *Proc. 8th National Computer Security Conf.*, 1985.
28. Nat'l Inst. Standards and Technology, *Proc. National Information Systems Security Conf.* (NISSC), 2001; http://csrc.nist.gov/nissc/.

29. Jelen, "Information Security: An Elusive Goal," pp. II-77–II-87.
30. G. Nibaldi, "Proposed Technical Evaluation Criteria for Trusted Computer Systems," MITRE, 25 Oct. 1979; www.cse.psu.edu/~tjaeger/cse544-s13/docs/niba79.pdf.
31. DoD Computer Security Center, "Trusted Computer System Evaluation Criteria," 1982.
32. DoD, "Trusted Computer System Evaluation Criteria: Final Draft," 1983.
33. DoD Computer Security Evaluation Center, "DoD Trusted Computer System Evaluation Criteria: Draft," 1982.
34. R.R. Schell, "Trusted Computer System Technical Evaluation Criteria," *Proc. 5th Seminar on the DoD Computer Security Initiative*, 1982, pp. 7–20.
35. Nibaldi, "Proposed Technical Evaluation Criteria for Trusted Computer Systems," p. 24
36. DoD Computer Security Center, "Trusted Computer System Evaluation Criteria," 1982, p. 24.
37. DoD Computer Security Evaluation Center, "DoD Trusted Computer System Evaluation Criteria: Draft," 1982, p. 13.
38. DoD, "Trusted Computer System Evaluation Criteria: Final Draft," p. 20.
39. DoD Computer Security Center, "Department of Defense Trusted Computer System Evaluation Criteria," 1983.
40. DoD Computer Security Center, "Department of Defense Trusted Computer System Evaluation Criteria," 1983, p. ii.
41. DoD Computer Security Initiative, *Proc. 6th Seminar on the DoD Computer Security Initiative*, Nat'l Bureau of Standards, 1983.
42. DoD Computer Security Center, "TPEP Historical EPL," 26 Aug. 1988; http://webapp1.dlib.indiana.edu/virtual_disk_library/index.cgi/1347159/FID1806/epl/historical.html.
43. DoD Computer Security Center, "Department of Defense Trusted Computer System Evaluation Criteria," 1983, pp. 15, 22.
44. DoD Computer Security Center, Dockmaster Criteria_Discussion forum, Univ. of Minnesota, Charles Babbage Inst. Document dates range from July 1985 to February 2000.
45. J. Epstein, interview with S.B. Lipner, 26–27 Feb. 2014.
46. S.B. Lipner, T. Jaeger, and M.E. Zurko, "Lessons from VAX/SVSfor High-Assurance VM Systems," *IEEE Security and Privacy*, vol. 10, no. 6, 2012, pp. 26–35.
47. D. MacKenzie and G. Pottinger, "Mathematics, Technology, and Trust: Formal Verification, Computer Security, and the U.S. Military," *IEEE Annals of the History of Computing*, vol. 19, no. 3, 1997, pp. 41–59.
48. V. Gligor et al., "On the Design and the Implementation of Secure Xenix Workstations," *Proc. IEEE Symp. Security and Privacy*, 1986, pp. 102–117.
49. D.C. Latham, "National Policy on Controlled Access Protection," Nat'l Telecomm. and Information Systems Security Committee, 15 July 1978; http://niatec.info/GetFile.aspx?pid=593.
50. S.B. Lipner, "Non-Discretionary Controls for Commercial Applications," *Proc. IEEE Symp. Security and Privacy*, 1982, pp. 2–10.
51. C.E. Landwehr, C.L. Heitmeyer, and J. McLean, "A Security Model for Military Message Systems," *ACM Trans. Computer Systems*, vol. 2, no. 3, 1984, pp. 198–222.
52. J. McLean, "Reasoning About Security Models," *Proc. IEEE Symp. Security and Privacy*, 1987, pp. 123–133.
53. Oracle, "Using Oracle Solaris 10 to Overcome Security Challenges," Aug. 2010; www.oracle.com/technetwork/server-storage/solaris10/solaris-10-security-167783.pdf.
54. Red Hat, "Red Hat Achieves Top Security Certification for Red Hat Enterprise Linux 6," 29 Oct. 2012; www.redhat.com/en/about/press-releases/red-hat-achieves-top-security-certification-for-red-hat-enterprise-linux-6.
55. D.J. Martin, "CCDB Work Items and Development Approach," *Proc. 9th Int'l Common Criteria Conf.*, 2008; www.commoncriteriaportal.org/iccc/9iccc/pdf/A2301.pdf.
56. Nat'l Information Assurance Partnership, "Position Statement Regarding the CC Evaluation of General Purpose Operating Systems," 17 Mar. 2014; https://www.niap-ccevs.org/Documents_and_Guidance/ccevs/GPOS%20Position%20Statement.pdf

Steven B. Lipner recently retired from Microsoft, where he was the Partner Director of Program Management in Trustworthy Computing Security. Lipner is the creator and was the long-time leader of Microsoft's Security Development Lifecycle (SDL) team that defines the SDL, develops associated tools and processes, and executes Microsoft's internal SDL process company wide. He was also a director and board chair of SAFECode, a nonprofit industry association dedicated to improving the security and integrity of software. He is a member of the ACM and IEEE Computer Society. Lipner has an MS in civil engineering from the Massachusetts Institute of Technology, and he attended the Harvard Business School's Program for Management Development. Contact him at lipner@outlook.com.