The Malware Rating System (MRS)™

Robert J. Bagnall and Geoffrey French
Veridian®
10455 White Granite Drive
Oakton, Virginia 22124

Abstract

The Malware Rating System (MRS) was created as an accurate, objective, and easy-to-use means of determining the threat from malicious code. Developed by both government and industry Information Security professionals the MRS is designed to impartially test malware. The system was designed to avoid the flaws found in other virus-rating architectures in several ways.

First, the MRS rates real malware, not malware creation kits or remote administration tools. In today's environment, Trojan horses, logic bombs, and other malicious logic also often behave with virus and worm-like characteristics. The MRS considers this convergent parasitic behavior and only rates these types of convergent code as well as the standard single-function types (virus, worm, logic bomb, and Trojan). Non-standard malicious logic that does not exhibit convergent parasitic behavior therefore is not rated by this scale. This ensures that the MRS analysis tables provide a more accurate, less subjective rating model. Second, the Malware Rating System is built and maintained by a team of industry and government professionals who hold no financial stake in the outcome of the rating issued. This objectivity is key to both how the program is received as well as its overall success. Third, the MRS is a category system readily understood by non-technical personnel. Based on both a mathematical model and comparison to similar code, the system uses a simple yet effective series of tables to feed a mathematical formula. The result is an Initial Category Rating (ICR) of 1 to 5 that reads like the Saffir-Simpson Hurricane Scale. The ICR is then issued to the community.

Finally, the MRS includes a regional factor (r-factor), a unique variable that allows adjustments to the initial rating by local analysts based upon how the code affects their individual operational capability. All of these elements make the Malware Rating System a more accurate and flexible basis for the evaluation of malware.

Background

The U.S. Department of Defense Computer Emergency Response Team (DoD-CERT), the Federal Computer Incident Response Center (FedCIRC), and the National Infrastructure Protection Center (NIPC) were established to serve as the U.S. government's focal points for threat assessment, warning, investigation, and response for threats or attacks against critical infrastructures. One of the missions they share in particular is to provide warnings of threats such as outbreaks of new malicious code, or malware. Providing this service to the community on a timely, accurate, and consistent basis enables these organizations to fulfill their missions as the key providers of American defense and infrastructure oversight. They often find themselves, however, speaking in different voices.

Discussions with government analysts that specialize in assessments of malicious code revealed several shortcomings in the way that government entities rate malicious code. First, reports often vary. Although there are good reasons why analysts from different organizations would come to different conclusions about the danger posed by the same malware, there should be consensus about the abilities of the code and its effects. Simultaneously, government agencies have had difficulty describing the threat of a single piece of malicious code to different systems. A large threat to Linux-based networks may pose no threat to a Windows NT network, for example. Finally, some agencies rate different phenomena (such as vulnerabilities in a Domain Name Server service) with the same system, stretching terms and rating criteria to force a fit.

Vendor-based rating systems have several of the same problems. Vendors tend to include many remote administration tools under the Trojan family (such as NetBus or Asmodeus), but exclude identical tools such as Microsoft’s SMS and Symantec’s PC Anywhere. Vendor systems also use variable scales, so that viruses considered of high concern today may be considered of low concern later, based upon the vendor’s patch update. Relative judgments over time cannot be used to provide an objective rating because they assume the use of a particular vendor’s anti-virus product, which may or may not be the best choice for the subject network. Vendor systems could easily be construed as having a fiscal interest in hostile code evaluation for the sake of profit as much as commitment to the community.

The Malware Rating System (MRS) represents an attempt to overcome some of these difficulties. Developed by both government and industry Information Security professionals, the MRS is designed to impartially test malware. The basis for the system and its method of implementation make the MRS an accurate and flexible model for the evaluation of malware.

Basis of the Malware Rating System

The MRS addresses the inconsistencies between and within previous rating systems in a number of ways. First, it focuses as much as possible on the code itself, limiting other factors that are difficult if not impossible to measure. Second, the aspects of the code are assigned scores based on defined thresholds. Third, these scores are weighted to produce a rating that reflects the priorities of experienced analysts. Fourth, the rating is communicated to the public through easily understood categories. Finally, the score can be adjusted by individual users or groups so that it reflects the way the code will affect a host’s specific vulnerabilities and characteristics.

Code-Focused Approach

As mentioned above, one of the problems that leads to inconsistencies in reporting on malicious code is a lack of precision regarding what is rated. The MRS defines some newer mobile code as a Parasite. That is, *"hostile code that exhibits convergent, parasitic behavior, whether accidentally or purposely released, and excluding hostile code creation kits and remote administration tools"*. This definition is the baseline by which all malicious logic is prescreened for acceptance prior to being rated by the Malware Rating System. As previously noted, convergent, parasitic behavior is defined as *malicious logic which exhibits any combination of the classic behavioral characteristics of viruses, worms, Trojans, or logic bombs.* The MRS rates both classic single-function malware (a virus, worm, Trojan, etc.) and parasites. As the characteristics and abilities of various types of malware converge, it becomes less useful to differentiate between them; hence, the broader term "parasite." The MRS excludes hostile code creation kits and remote administration tools.

Another characteristic of a rating system designed for consistent reporting is that it must be based as much as possible on objective measures. For this reason, the MRS focuses on the code itself, that is, its potential in an unprotected environment and its qualities as compared to other similar hostile code. It eliminates such factors as the number of worldwide infections, which is difficult to measure reliably. Instead, it focuses on three specific aspects of the code: (1) its payload, (2) its potential to proliferate, and (3) its hostility (see Table 1).

Table 1: Criteria for Rating Malware

Criterion	Description
Payload potential	Potential of the code module to degrade or damage its target
Proliferation potential	How fast or easily the code is passed along
Hostility level	The intent behind the payload (i.e., whether it is designed to degrade, destroy, or merely annoy)

Threshold-Based Scores

The second characteristic of the MRS that aids in consistent reporting is its reliance on thresholds to rate the individual criteria. Both the payload and the proliferation potential are rated on a scale of one to ten (with ten indicating most destructive or most easily proliferated). The malicious code's hostility level is rated between one and five. Although other systems also employ numbered ratings, the MRS proposes thresholds for these scores. If these thresholds are commonly accepted, this should eliminate differences in the scores between agencies. Tables 2 through 4 describe the proposed thresholds.

Table 2: Proposed Thresholds that Determine the Payload Rating

Rating	Description	Example
10	Polymorphic -- patch immune, or previously undefined	
9	Deletes essential files, infects network; can crash a network with a bandwidth capacity overflow	Chernobyl
8	Deletes, modifies, or overwrites essential files and infects network	New Love, CIH
7	Modifies or overwrites non-essential files, infects network; can crash a network with a bandwidth capacity overflow	Pretty Park, I Love You
6	Deletes non-essential files, infects network	I Love You
5	Deletes non-essential files	
4	Can crash a network with a bandwidth capacity overflow	Melissa
3	Infects essential files	Boot-437
2	Infects non-essential files	
1	No lasting effects (only self replicates)	Bubbleboy, Happy-99

Table 3: Proposed Thresholds that Determine the Proliferation Rating

Rating	Description	Example
10	Network spread (via e-mail, IRC, HTML, etc), does not need attachment to be opened to replicate; not OS specific	
9	Network spread (via e-mail, IRC, HTML, etc), does not need attachment to be opened to replicate; OS specific	
8	Network spread (via e-mail, IRC, HTML, etc), does not need attachment to be opened to replicate; application specific	Melissa, Bubbleboy
7	Network spread (via e-mail, IRC, HTML, etc), needs attachment to be opened to replicate; OS specific	Chernobyl, Pretty Park, Happy-99, CIH
6	Network spread (via e-mail, IRC, HTML, etc), needs attachment to be opened to replicate; application specific	I Love You, New Love,
5	File spread; not OS specific	
4	File spread; OS specific	Boot-437
3	File spread; application specific	
2	Planted; not OS specific	
1	Planted; OS specific	

Table 4: Proposed Thresholds that Determine the Hostility Rating

Rating	Description	Example
5	Meant to be destructive and disruptive world wide	Pretty Park, I Love You
4	Meant to be destructive	Chernobyl
3	Meant to be disruptive world wide	Bubbleboy
2	Meant to be locally disruptive	Boot-437
1	Accidental release	Morris

Weighted Formula

In addition to the consistent determination of the individual criteria, the MRS combines the scores in a manner that reflects the values of experienced analysts. By taking into account the characteristics of the code that analysts and systems administrators consider to be the most important (i.e., that cause malware to pose the most serious threat), the MRS provides a carefully considered rating on a scale of one to 100.

To make the system as objective as possible, the two primary criteria are the code's payload and its potential to proliferate. The payload potential (i.e., its effects on a host or network) is the basis for the score because it represents the consequences of infection. For this reason, the individual payload score can contribute up to three-quarters of the rating (that is, it is multiplied by 7.5). This ensures that the payload plays a significant role in the code's final categorization (see the following section) but cannot determine that a code attains the highest rating category without an above average degree of proliferation.

Input from analysts indicated that a code's proliferation potential influences the degree to which the code threatens systems in general. Therefore, the proliferation score can have a positive effect (if a malicious code is very easily spread from one system to another) or a negative effect (if a particular malicious code needs very specific circumstances to proliferate). To represent this mathematically, the individual score of the proliferation potential is multiplied by 5 and then subtracted from 25. This total is then subtracted from the payload's score. In this way, the proliferation potential can increase or decrease a code's category by two full rankings (see the following section).

The final criterion is hostility, that is, the type of effect that the code is intended to have on a network. Unlike the proliferation potential, the hostility intent of a certain piece of malicious code cannot reduce its score. Although a piece of malware that is designed to target a specific type of network may be of higher concern than one that is not, one that destroys or degrades a network or its data despite the intentions of its designer is of no less concern because of the intent. This factor helps distinguish between code released accidentally and that released purposefully to do harm. The score for the intent behind the code (its "hostility") is simply added to the composite score so that a code on the cusp of a higher ranking can increase its ranking in situations where the intent might influence the threat it poses. The result is a score between one and

100. (Note that it is theoretically possible to have some scores below zero and some above 100. These are rounded to the margins of the scale.)

Initial Category Rating (ICR)

In order to provide a truly effective rating system, it was decided that the MRS should translate the calculated rating into a form that could be easily recognized and understood by even the most novice computer users. A category system, similar to a hurricane intensity scale, was selected as a way to best provide this functionality. It offers the local analyst or system administrator, as well as even the novice Internet user, the ability to understand the severity of a particular piece of malware based upon easy-to-understand, accurate, and objective means.

The final step in the rating process, therefore, is the assignment of a warning category. After the individual criteria of the code have been rated and combined, the calculated rating is then simplified by a corresponding category (see Table 5). The categories are intended to give a simple description to the destructive potential of the code (e.g., minimal or dangerous). In this way, the Malware Rating System makes the rating and the process easy to understand and therefore accessible for users of all levels expertise. Examples are presented in Table 6.

Table 5: Category Ratings for Malware

Score	Category	Danger Description
0 – 20	1	Minimal
21 – 40	2	Low
41 – 60	3	Dangerous
61 – 80	4	Extreme
81 – 100	5	Catastrophic

Table 6: Examples of Ratings for Malware Payload Proliferation Hostility

Name	Family	Payload Potential	Proliferation Potential	Hostility Level	Score[a]	Category[b]	Notes
Melissa virus	Macro	4	8.5	3	50.5	3	Although it contains a basically harmless payload, Melissa spreads very efficiently, degrading network operations.
Boot-437 (a.k.a. “Bath”)	Boot sector	3	4	2	19.5	1	Boot-437 does not contain a dangerous payload; it only replicates from floppy disk, minimizing its effects.
CIH	Executable	8	7	5	75	4	The highly dangerous payload and relatively easy proliferation make CIH an ideal destructive tool.
NewApt	Worm	4	10	4	59	3	NewApt acts much like Melissa in how it replicates, yet is designed to alter system files as well. This leads to its slightly higher score.
W32-Park (a.k.a. “Pretty Park”)	Worm	7	7	5	67.5	4	Pretty Park uses e-mail replication, alters system files, and opens IRC channels to give itself Trojan-like capability for a remote intruder.

Notes:

[a] Score is based on the formula (7.5P - [25-5Pr] + h) where P = payload potential, Pr = proliferation potential, and h = hostility level.

[b] The score is translated to a category as described in Table 2.

The r-Factor

Although MRS delivers a standard and consistent rating for a specific piece of malicious code, it does not attempt to describe the risk to every system. Instead, the regional factor, or “r-factor”, variable is the final determinate in how potentially dangerous a piece of malware can be to a specific enterprise. The r-factor determines the effects of the user’s systems, security, and local protection measures on the potential of the code itself. Although the MRS category rating will remain the same (to support future category ratings of similar hostile code as it is developed and released), the r-factor makes it possible to provide the most realistic and timely assessment of the effect of hostile code as it applies to your system or network.

The r-factor allows local network defense teams, who are most familiar with a given network, the ability to reassess the destructive potential of hostile code by using the initial national-level MRS assessment as the starting point. For example, if a variant of the Melissa virus were detected, an agency's information security officer could receive a report that the virus has been rated as a category 3 threat (i.e., dangerous). Taking into account that his or her network is Linux-based and does not use macros, the officer uses the r-factor to reassess the virus' threat to be a category 1 threat (i.e., minimal).

Table 7: Proposed Thresholds that Determine the r-Factor

Rating	Description	Example
-2	Uses different OS than your network	Ramen worm on a Windows network
-1	Uses different application than your network (note you can still be affected by receiving a large number of e-mail messages)	Melissa on a Linux network
-1	Patch available and applied	
+1	Affects bandwidth	
+1	Affects storage or buffer space	
+2	Targeted for your type of network	Melissa/I Love You against an Outlook & Exchange mail system
+4	Targeted for files with specific information that you might have or targeted for your IP; Polymorphic or patch immune	I Love You against jpegs or Ramen against an unpatched Linux network

Implementation and Review

The Malware Rating System is designed for a top-down implementation, taking advantage of the current government hierarchy. The MRS is a national-level asset to aid the malware defense effort of the three top government oversight bodies; DoD-CERT, NIPC, and FED-CIRC. Because these organizations oversee the infrastructure defense effort for the DoD, commercial, and federal government communities respectively, the MRS provides a single voice coordination capability that guides the entire American infrastructure defense effort as a whole.

Implementing the MRS at this level provides three unique solutions not currently available. First, it eliminates the confusion of three defense sectors offering conflicting ratings and statements during a malware crisis. Second, it provides a single voice of guidance from the three lead organizations responsible for protection and response coordination and oversight. Third, the MRS relies upon local defense teams (such as AFCERT, DOE-CERT, or USPS-CERT), involving them in the threat decision-making and protection response process from the outset of the initial category rating.

The MRS will be automated and given a graphical user interface so as to be most accessible to the DoD, federal government, and commercial communities. Responsible parties designated to deal with malicious code concerns within DoD-CERT, FED-CIRC, and NIPC will use an XML form that will pose a series of questions about the next new piece of hostile code. The interface will then parse the answers, calculate the initial category rating based upon the mathematical formula, and post the results. When posted, the MRS will allow local analysts to compute their individual r-factors and produce a report.

The Malware Rating System will be reviewed annually for accuracy and evolutionary improvement by the participating members of the Malware Defense Forum (MDF). This forum will include the NIPC, DoD-CERT, and FedCIRC as the three founding members and will include associate membership by any government or commercial entity who maintains a participating membership or whose defense posture is guided by one of the three founding members. The purpose of the forum's review of the MRS is to evaluate the currency and validity of the tables and formula data that produce the ICRs. By ensuring that the definitions within the tables remain relevant the forum keeps the MRS current with the evolutionary developments of the hostile code environment. The MDF also ensures that the MRS is a community effort, driven by all the contributing members and not simply a small minority.

What the Future Holds

Were the MRS implemented as is today, the quest for a viable malware and mobile code rating system would still be in its infancy. Evolutionary concerns for the MRS include addressing its current lack of vulnerability rating capability. As the Information Assurance world continues to merge, the assessment of malware as part of the vulnerability assessment and rating process must also merge. This should take into account, at a minimum, vulnerabilities related to software specifically as they are most closely related to the malware that will exploit them over time. Additionally, the emerging presence of agent-based malware, that is -- that code which can assess and exploit system vulnerabilities through a multi-step process to the root or administrator level without the guidance or control of a human operator, must be addressed in a future version of the MRS. These and other not yet realized concerns will be merged into future versions of the Malware Rating System.

Conclusion

To accurately assess current and future dangers to government and commercial network operations, it is vital to have a mechanism whereby data can be analyzed accurately and succinctly, without bias or financial interest. MRS was developed to enable the end-user analysts to assess malware and compare its effect against their system or network. Enabling a web interface across the government and commercial communities provides an instantly updated assessment of malicious code for all industry, civilian or government.

But regardless of the technology behind its perpetuation, the Malware Rating System represents an idea long overdue in government: the provision for a unified, focused, accurate, and logical approach to rating mobile code and malware. However the MRS evolves toward a final, accepted standard, the concept itself remains a standard that must be implemented for the American Infrastructure Protection community to remain viable against all aggressors.